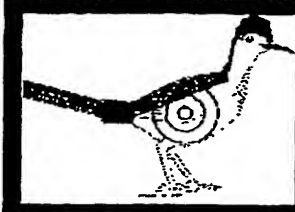


FIG. 1


1/40

File Edit View Go Communicator

Bookmarks Location: <http://victoria.inpharmatica.co.uk/~volker/BPD3target.html> WIPD



Target Mining Interface



### Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA):  and chain (e.g. B):
- OR
- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

### Select Database

Release:

### Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first\_PB\_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

Filter for the following SPECIES:

<input type="checkbox"/> Homo sapiens	<input type="checkbox"/> Rattus norvegicus (Rat)	<input type="checkbox"/> Mus musculus (Mouse)	<input type="checkbox"/> Danio rerio (Zebra fish)
---------------------------------------	--	---	---

100%

FIG. 2A

2/40

File Edit View Go Communicator Help													
Bookmarks Location http://London-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl													
2) 84 additional hits identified by both, Genome Threader and PSI-BLAST:													
Combined Genome Threader and PSI-Blast output: PSI-BLAST values are shown in maroon!													
Add2list	EPD links	WWW link	Title	Organism	Div.	%ID (GT, PSI)	Query rgn. (GT, PSI)	Target rgn. (GT, PSI)	Aln. score (GT)	Conf. (GT)	1st iter. (PSI)	Best iter. (PSI)	Ex- v (P)
J	<a href="#">AAAS9544.1</a> <a href="#">drill through</a> <a href="#">Top50BlastHits</a> <a href="#">Red Seq. View</a>	<a href="#">AAAS9544.1</a>	Not given	Homo sapiens	PRI	33.7%, 33% unmaskedSW	4-183, 4-183	150-336, 150-336	449	100% unmaskedGT	1	2	3E
J	<a href="#">AAB24821.1</a> <a href="#">drill through</a> <a href="#">Top50BlastHits</a> <a href="#">Red Seq. View</a>	<a href="#">AAB24821.1</a>	leukocyte integrin alpha chain	Homo sapiens	PRI	33.7%, 33% unmaskedSW	4-183, 4-183	150-336, 150-336	449	100% unmaskedGT	1	2	3E
J	<a href="#">Q99715</a> <a href="#">drill through</a> <a href="#">Top50BlastHits</a> <a href="#">Red Seq. View</a>	<a href="#">Q99715</a>	COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR.	Homo sapiens (Human).	PRI	28.8%, 26% unmaskedSW	4-181, 2-174	140-318, 2321-2495	440	100% unmaskedGT	1	3	2E
J	<a href="#">P20701</a> <a href="#">drill through</a> <a href="#">Top50BlastHits</a> <a href="#">Red Seq. View</a>	<a href="#">P20701</a>	LEUKOCYTE ADHESION GLYCOPROTEIN LFA-1 ALPHA CHAIN PRECURSOR (LEUKOCYTE FUNCTION ASSOCIATED MOLECULE 1, ALPHA CHAIN) (CD11A) (INTEGRIN ALPHA-L).	Homo sapiens (Human).	PRI	100%, 100% unmaskedSW	1-183, 1-183	153-335, 153-335	423	100% unmaskedGT	1	1	1E
J	<a href="#">AAC31672.1</a> <a href="#">drill through</a> <a href="#">Top50BlastHits</a> <a href="#">Red Seq. View</a>	<a href="#">AAC31672.1</a>	leukocyte function-associated molecule-1 alpha subunit	Homo sapiens	PRI	99.5%, 99% unmaskedSW	1-183, 1-183	153-335, 153-335	423	100% unmaskedGT	1	1	1E
J	<a href="#">CAA72402.1</a> <a href="#">drill through</a> <a href="#">Top50BlastHits</a> <a href="#">Red Seq. View</a>	<a href="#">CAA72402.1</a>	collagen type XIV	Homo sapiens	PRI	29.1%, 29% unmaskedSW	2-180, 2-180	5-185, 5-185	422	100% unmaskedGT	1	2	2E
J	<a href="#">AAB38702.1</a> <a href="#">drill through</a> <a href="#">Top50BlastHits</a> <a href="#">Red Seq. View</a>	<a href="#">AAB38702.1</a>	cartilage matrix protein	Homo sapiens	PRI	31.7%, 27% unmaskedSW	4-183, 2-182	275-455, 39-223	413	100% unmaskedGT	1	2	2E
J	<a href="#">CAB70853.1</a> <a href="#">drill through</a> <a href="#">Top50BlastHits</a> <a href="#">Red Seq. View</a>	<a href="#">CAB70853.1</a>	hypothetical protein	Homo sapiens	PRI	28%, 28% unmaskedSW	1-180, 1-183	437-620, 437-624	406	100% unmaskedGT	1	2	1E
J	<a href="#">CAA27972.1</a> <a href="#">drill through</a> <a href="#">Top50BlastHits</a> <a href="#">Red Seq. View</a>	<a href="#">CAA27972.1</a>	Not given	Homo sapiens	PRI	20.5%, 20% unmaskedSW	3-183, 2-181	1497-1673, 1689-1873	405	100% unmaskedGT	2	3	1E
J	<a href="#">AAB59512.1</a> <a href="#">drill through</a> <a href="#">Top50BlastHits</a> <a href="#">Red Seq. View</a>	<a href="#">AAB59512.1</a>	Not given	Homo sapiens	PRI	20.5%, 20% unmaskedSW	3-183, 2-181	758-934, 950-1134	405	100% unmaskedGT	2	3	1E
J	<a href="#">CAA07569.1</a> <a href="#">drill through</a> <a href="#">Top50BlastHits</a> <a href="#">Red Seq. View</a>	<a href="#">CAA07569.1</a>	matrin-4	Homo sapiens	PRI	28.1%, 25% unmaskedSW	1-183, 1-183	342-528, 31-217	403	100% unmaskedGT	1	2	2E

3/40

FIG. 2B

1) 509 hits identified by Genome Threader only:

Addl list	Redundant sequence display	BPD link	WWW link	Title	Organism	Div.	%ID	Query rgn.	Target rgn.	Aln. score	Conf.
↓	<a href="#">Red. Seq. View</a>	<a href="#">AAF71133.1 drill through Top50BlastHits</a>	<a href="#">AAF71133.1</a>	PRO2789	Homo sapiens	PRI	13.8% unmaskedSW	109-179	1-80	122	100% unmaskedGI
↓	<a href="#">Red. Seq. View</a>	<a href="#">CAB52192.1 drill through Top50BlastHits</a>	<a href="#">CAB52192.1</a>	G7c protein	Homo sapiens	PRI	9.6% unmaskedSW	7-117	20-124	82	99.18% unmaskedGI
↓	<a href="#">Red. Seq. View</a>	<a href="#">CAA82910.1 drill through Top50BlastHits</a>	<a href="#">CAA82910.1</a>	basic transcription factor 2, 44 kD subunit	Homo sapiens	PRI	11% unmaskedSW	5-164	61-225	75	99.18% unmaskedGI
↓	<a href="#">Red. Seq. View</a>	<a href="#">AAD21820.1 drill through Top50BlastHits</a>	<a href="#">AAD21820.1</a>	NG37	Homo sapiens	PRI	9.6% unmaskedSW	7-117	318-422	82	98.86% unmaskedGI
↓	<a href="#">Red. Seq. View</a>	<a href="#">BAA20781.1 drill through Top50BlastHits</a>	<a href="#">BAA20781.1</a>	Not given	Homo sapiens	PRI	16.4% unmaskedSW	5-114	1836-1950	78	98.52% unmaskedGI
↓	<a href="#">Red. Seq. View</a>	<a href="#">AAA36154.1 drill through Top50BlastHits</a>	<a href="#">AAA36154.1</a>	Not given	Homo sapiens	PRI	13.2% unmaskedSW	5-112	10-137	79	98.16% unmaskedGI
↓	<a href="#">Red. Seq. View</a>	<a href="#">AAF03046.1 drill through Top50BlastHits</a>	<a href="#">AAF03046.1</a>	candidate tumor suppressor protein DICE1	Homo sapiens	PRI	13.8% unmaskedSW	5-113	4-131	79	97.07% unmaskedGI
↑	<a href="#">Red. Seq. View</a>	<a href="#">AAC74654.1 drill through Top50BlastHits</a>	<a href="#">AAC74654.1</a>	orf, hypothetical protein	Escherichia coli	BCT	16.7% unmaskedSW	5-89	250-335	78	95.09% unmaskedGI
↓	<a href="#">Red. Seq. View</a>	<a href="#">AAB60942.1 drill through Top50BlastHits</a>	<a href="#">AAB60942.1</a>	breast cancer suppressor candidate 1	Homo sapiens	PRI	16.2% unmaskedSW	63-156	2-90	80	93.33% unmaskedGI
↓	<a href="#">Red. Seq. View</a>	<a href="#">AA67537.1 drill through Top50BlastHits</a>	<a href="#">AA67537.1</a>	glycoprotein IIIa	Homo sapiens	PRI	18.9% unmaskedSW	3-112	111-251	72	92.08% unmaskedGI reverse Hit
↓	<a href="#">Red. Seq. View</a>	<a href="#">AAA52589.1 drill through Top50BlastHits</a>	<a href="#">AAA52589.1</a>	Not given	Homo sapiens	PRI	19.6% unmaskedSW	3-112	138-277	74	91.34% unmaskedGI reverse Hit
↓	<a href="#">Red. Seq. View</a>	<a href="#">AAB71380.1 drill through Top50BlastHits</a>	<a href="#">AAB71380.1</a>	platelet membrane glycoprotein IIIa beta subunit	Homo sapiens	PRI	19.6% unmaskedSW	3-112	138-277	74	91.34% unmaskedGI reverse Hit
↓	<a href="#">Red. Seq. View</a>	<a href="#">AAA35927.1 drill through Top50BlastHits</a>	<a href="#">AAA35927.1</a>	Not given	Homo sapiens	PRI	19.6% unmaskedSW	3-112	138-277	74	91.34% unmaskedGI

File Edit View Go Communicator Help

Bookmarks  Go to: <http://london-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl>

630 out of these 64S PSI-BLAST matches were identified using 'positive iterations':

[illegible]

15 out of these 645 PSI-BLAST matches were identified using 'negative iterations':

[illegible]

### A2) Genome Threader Matches:

27229 matches found by Genome Threading:

[illegible]

5/40

**inpharmatica**

**Redundant Sequence Display**

File Edit Search Type Help

Contains:  
2 Sub-sequences.  
0 PROSITE hits.  
0 PRINTS hits.

Representative code: BAA15585.1 Length: 427 Organism name: *Escherichia coli* [...additional annotation]

Aligned sequences

Code
BAA15585.1
P76235
AAC74854.1

Primary database information

GenBank	
protein_id	Details
BAA15585.1	Tax ID: 562
	Organism name: <i>Escherichia coli</i>
	EC Number: Not available
	Gene Name: yzdC
	Description: Not given
	Other links:
	GI: 1736412 TAXON: 562

SWISS-PROT

Code	
Details	
P76235	Tax ID: 562
	Organism name: <i>Escherichia coli</i>
	Gene Name: YEAH
	EC Number: Not available
	Description: HYPOTHETICAL 49.4 KDA PROTEIN IN GAPA-RND INTERGENIC REGION.

Tools

Sequences

Lipids

Start

inpharmatica

EPD4-FEI.4

inpharmatica

Fagan, Richa...

Inbox - Outb...

Target Minin...

Microsoft Po...


FIG. 3

6/40  
FIG. 4


File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Stop

Bookmarks Location: <http://www.sanger.ac.uk/cgi-bin/Pfam/nph-search.cgi>



**Pfam**  
Protein families database of alignments and HMMs  
Home | Keyword search | Protein search | DNA search | Browse Pfam | Taxonomy search | Help




### Results for gi|1788084|gb|AAC74854.1|

There were no matches to Pfam-A (including borderline matches) for gi|1788084|gb|AAC74854.1|

**Matches to Pfam-B**

Domain	Start	End	Evalue	Alignment
Pfam-B_39416	233	423	3.7e-103	Align

 [427 residues]

### Alignments of Pfam-B domains to best-matching Pfam-B sequence

Format for fetching alignments to Pfam-B families: ☐ Hypertext linked to swisspfam

Query gi|1788084|gb|AAC74854.1|233-423 matching Pfam-B\_39416

```

YEAH_ECOLI 233 DLRYKNYEKRPDPSSQAVMFCIMDVSGSMDQSTKDMAKRFYILLYLFLSR 282
gi|1788084|gb|AAC74854.1| 233 DLRYKNYEKRPDPSSQAVMFCIMDVSGSMDQSTKDMAKRFYILLYLFLSR 282

YEAH_ECOLI 283 TYKNVEVVYIRNHTQAKEVDEHEFFYSQETGGTIVSSALKLMDEVVKERY 332
gi|1788084|gb|AAC74854.1| 283 TYKNVEVVYIRNHTQAKEVDEHEFFYSQETGGTIVSSALKLMDEVVKERY 332

YEAH_ECOLI 333 NPAQWHIYAAQASDGDNWADD SPLCHEILAKKLLPVVRYYSYIEITRRAH 382
gi|1788084|gb|AAC74854.1| 333 NPAQWHIYAAQASDGDNWADD SPLCHEILAKKLLPVVRYYSYIEITRRAH 382

YEAH_ECOLI 383 QTLWREYEHLSQSTFDNFMQHIRDQDDIYPVVFRELPHKQNA 423
gi|1788084|gb|AAC74854.1| 383 QTLWREYEHLSQSTFDNFMQHIRDQDDIYPVVFRELPHKQNA 423

```

If you think there is anything wrong with this script, please contact [Pfam](#)

100%



FIG. 5

7/40


File Edit View Go Communicator Help	
Bookmarks Location: <a href="http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&amp;db=">http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&amp;db=</a>	
LOCUS	AAC74854 427 aa BCT 01-DEC-2000
DEFINITION	orf, hypothetical protein [Escherichia coli K12].
ACCESSION	AAC74854
PID	g1788084
VERSION	AAC74854.1 GI:1788084
DBSOURCE	locus AE000273 accession AE000273.1
KEYWORDS	
SOURCE	Escherichia coli K12.
ORGANISM	Escherichia coli K12 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE	1 (residues 1 to 427)
AUTHORS	Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, M.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.
TITLE	The complete genome sequence of Escherichia coli K-12
JOURNAL	Science 277 (5331), 1453-1474 (1997)
MEDLINE	97426617
PUBMED	9278503
REFERENCE	2 (residues 1 to 427)
AUTHORS	Blattner, F.R.
TITLE	Direct Submission
JOURNAL	Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE	3 (residues 1 to 427)
AUTHORS	Blattner, F.R.
TITLE	Direct Submission
JOURNAL	Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE	4 (residues 1 to 427)
AUTHORS	Plunkett, G. III.
TITLE	Direct Submission
JOURNAL	Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT	This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible ( <a href="http://cgsc.biology.yale.edu">http://cgsc.biology.yale.edu</a> ). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site ( <a href="http://www.genetics.wisc.edu">http://www.genetics.wisc.edu</a> ). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names. Method: conceptual translation.
FEATURES	Location/Qualifiers
source	1..427 /organism="Escherichia coli K12" /strain="K12" /sub_strain="MG1655" /db_xref="taxon:83333"
Protein	1..427 /function="orf; Unknown" /product="orf, hypothetical protein"
CDS	1..427 /gene="yeam" /coded_by="1788078:6385..7668" /transl_table=11 /note="o427; This 427 aa ORF is 28 pct identical (43 gaps) to 327 residues of an approx. 312 aa protein YZDC_BAC5U SW: P45742"
ORIGIN	1 mtwfidrrln gknksmvnrq rllrrpkaqi kqsiseaink rsvtdvdsge svstiptedis 61 epmfhgqrgg lhrvhpgnd hfvndrier pggggggsgs gggqasqdg gqdeivqis 121 kdeyldllfe dlalpnlqm qqrqltepkt hragnptangv panisvvrsl qnslarrtam 181 taqkrrelha leenlaish sepaqllee rlrkeiaelr akiervpfid tfdlryknye

8/40

FIG. 6A

File Edit Search Type: Help

# Accession Code Query



Tools

- ←
- 
- ?
- ?
- ?
- X

Sequences

- ? ?
- ? ?
- ? ?
- ? Acc y123

Ligands

- ? RDP NRD FRD
- ? Phe ALA GLU

Select a database and enter a code:

Database	
<b>GenBank</b>	
Accession	<input checked="" type="radio"/>
Accession Version	<input type="radio"/>
GI-NID	<input type="radio"/>
protein_id	<input type="radio"/>
CEPID	<input type="radio"/>
<b>SWISS-PROT</b>	
Accession	<input type="radio"/>
ID	<input type="radio"/>
<b>PDB</b>	
ID	<input type="radio"/>

AAC74854.1

Submit Reset



9/40

File Edit Search Type Help

Aligned Sequence Display

Query details:  
AAC74854.1:orf, hypothetical protein

inpharmatica

Tools

Sequences

Ligands

Page 1 of 2

Total hits: 36  
Selected: 0

Total selected: 0

View alignment

Inpharmatica Genome Threader results

Cluster	Detail	Code	Title	Organism	%ID	Query rgn.	Target rgn.	Aln. score	Method	Confidence
1	1LFAA	CD11A	I-DOMAIN WITH BOUND MN++	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	78	Local	Marginal(95%)
2	1ZOOA	CD11A	I-DOMAIN WITH BOUND MAGNESI...	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	78	Local	Marginal(95%)
3	1ZOOB	CD11A	I-DOMAIN WITH BOUND MAGNESI...	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	78	Local	Marginal(95%)
4	1ZOPB	CD11A	I-DOMAIN WITH BOUND MAGNESI...	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	78	Local	Marginal(95%)
5	1ZOPA	CD11A	I-DOMAIN WITH BOUND MAGNESI...	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	78	Local	Marginal(95%)
6	1BHO1	MAC-1	I-DOMAIN MAGNESIUM COMPLEX	<i>H.sapiens</i>	15.1	250 - 365	4 - 128	76	Local	Low(94%)
7	1IDN1	MAC-1	I-DOMAIN METAL FREE	<i>H.sapiens</i>	15.1	250 - 365	4 - 128	76	Local	Low(94%)
8	1BHQ2	MAC-1	I-DOMAIN CADMIUM COMPLEX	<i>H.sapiens</i>	15.1	250 - 365	4 - 128	76	Local	Low(94%)
9	1BHQ1	MAC-1	I-DOMAIN CADMIUM COMPLEX	<i>H.sapiens</i>	15.1	250 - 365	4 - 128	76	Local	Low(94%)
10	1BHO2	MAC-1	I-DOMAIN MAGNESIUM COMPLEX	<i>H.sapiens</i>	15.1	250 - 365	4 - 128	76	Local	Low(94%)
11	1IDN2	MAC-1	I-DOMAIN METAL FREE	<i>H.sapiens</i>	15.1	250 - 365	4 - 128	76	Local	Low(94%)
12	1ATZB	HUMAN VON WILLEBRAND FACTOR A3 DO...		<i>H.sapiens</i>	15.1	244 - 352	1 - 126	70	Local	Low(94%)
13	1ATZA	HUMAN VON WILLEBRAND FACTOR A3 DO...		<i>H.sapiens</i>	15.3	250 - 335	5 - 101	72	Local	Low(93%)
14	1DGQA	NMR SOLUTION STRUCTURE OF THE INSR...		<i>H.sapiens</i>	15.3	250 - 331	8 - 99	75	Local	Low(93%)
15	1IDO	I-DOMAIN FROM INTEGRIN CR3, MG2+ BO...		<i>H.sapiens</i>	15.1	250 - 365	4 - 128	75	Local	Low(92%)
16	1JLM	I-DOMAIN FROM INTEGRIN CR3, MN2+ BO...		<i>H.sapiens</i>	15.1	250 - 365	4 - 128	75	Local	Low(89%)
17	1ZON	CD11A	I-DOMAIN WITHOUT BOUND CATI...	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	72	Local	Low(83%)
18	1LFA8	CD11A	I-DOMAIN WITH BOUND MN++	<i>H.sapiens</i>	18.0	250 - 333	5 - 97	66	Local	Low(80%)
19	1OAKA	CRYSTAL STRUCTURE OF THE VON WILLE...		<i>H.sapiens</i>	20.3	251 - 301	1 - 59	72	Local	Low(72%)
20	1AUQ	A1-DOMAIN OF VON WILLEBRAND FACTOR		<i>H.sapiens</i>	21.2	250 - 301	10 - 59	89	Local	Low(66%)
21	1A08A	A3-DOMAIN OF VON WILLEBRAND FACTOR		<i>H.sapiens</i>	15.2	245 - 331	2 - 99	63	Local	Low(53%)
22	1A08B	A3-DOMAIN OF VON WILLEBRAND FACTOR		<i>H.sapiens</i>	15.2	245 - 331	2 - 99	63	Local	Low(53%)
23	1A08C	I-DOMAIN FROM INTEGRIN ALPHA2-BETA1		<i>H.sapiens</i>	13.1	242 - 335	1 - 105	53	Local	Low(21%)
24	1RYPD	CRYSTAL STRUCTURE OF THE 20S-PROTE...		<i>Saccharomyces</i>	13.1	246 - 305	166 - 226	58	Local	Low(21%)
25	1RYPB	CRYSTAL STRUCTURE OF THE 20S-PROTE...		<i>Saccharomyces</i>	13.1	246 - 305	166 - 226	58	Local	Low(21%)
26	1CK4B	CRYSTAL STRUCTURE OF RAT A1B1 INTEG...		<i>Rattus norvegicus</i>	10.5	250 - 331	5 - 96	55	Local	Low(20%)

GenBank

SWISS-PROT

PDB

FIG. 6B

10/40

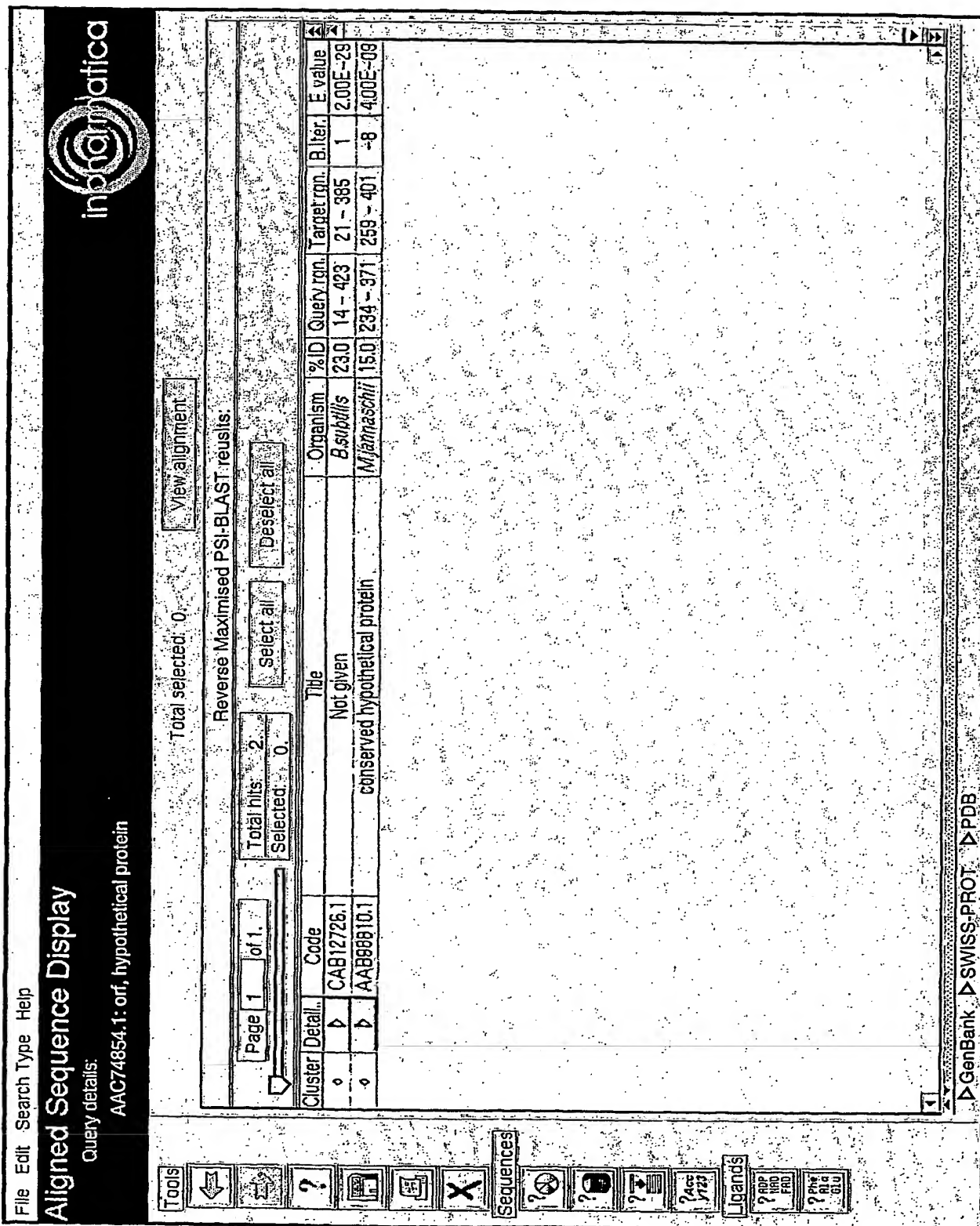


FIG. 6C

## FIG. 7

AlEye output (January 4, 2002 3:07 PM)

11/40

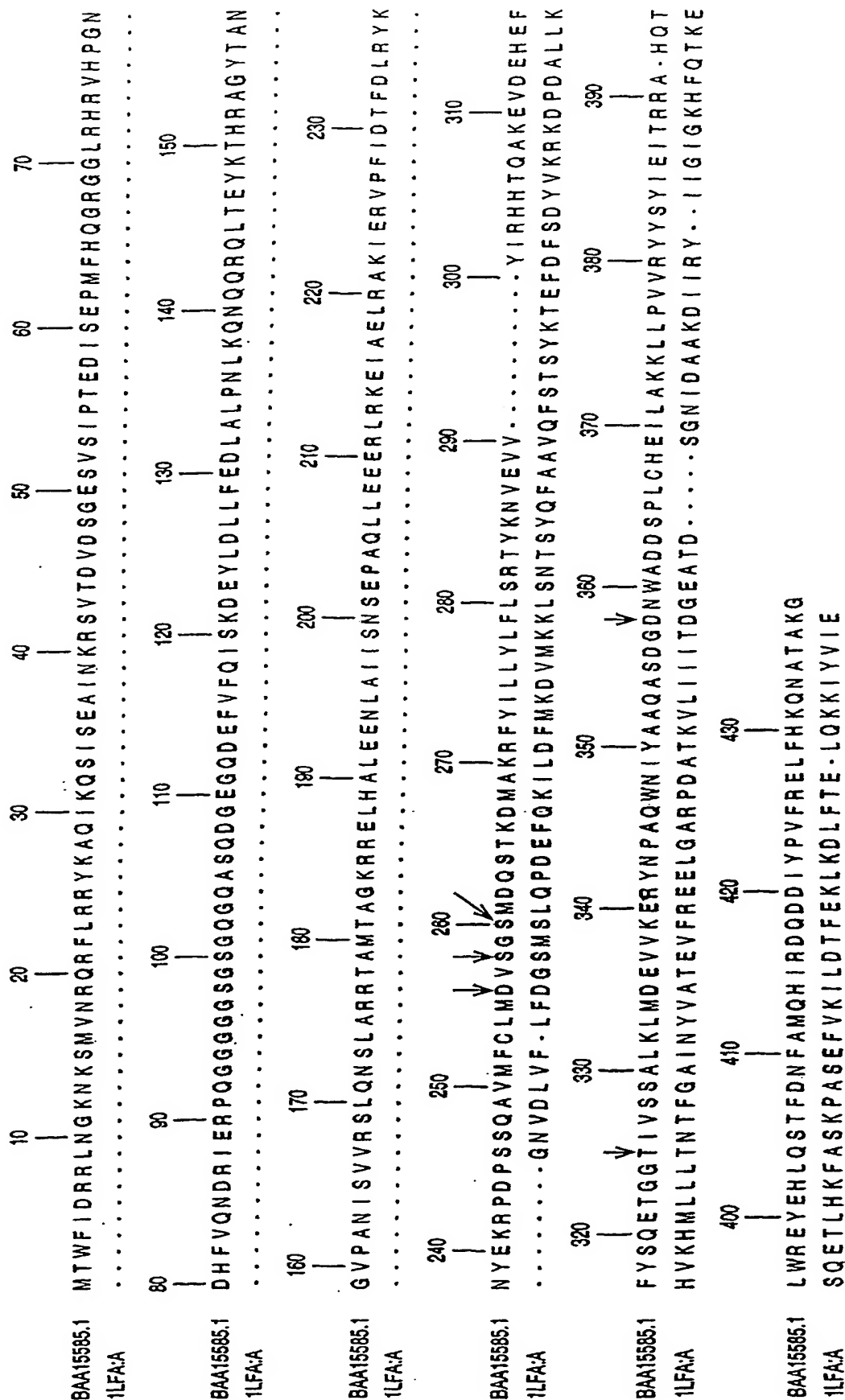


FIG. 8B

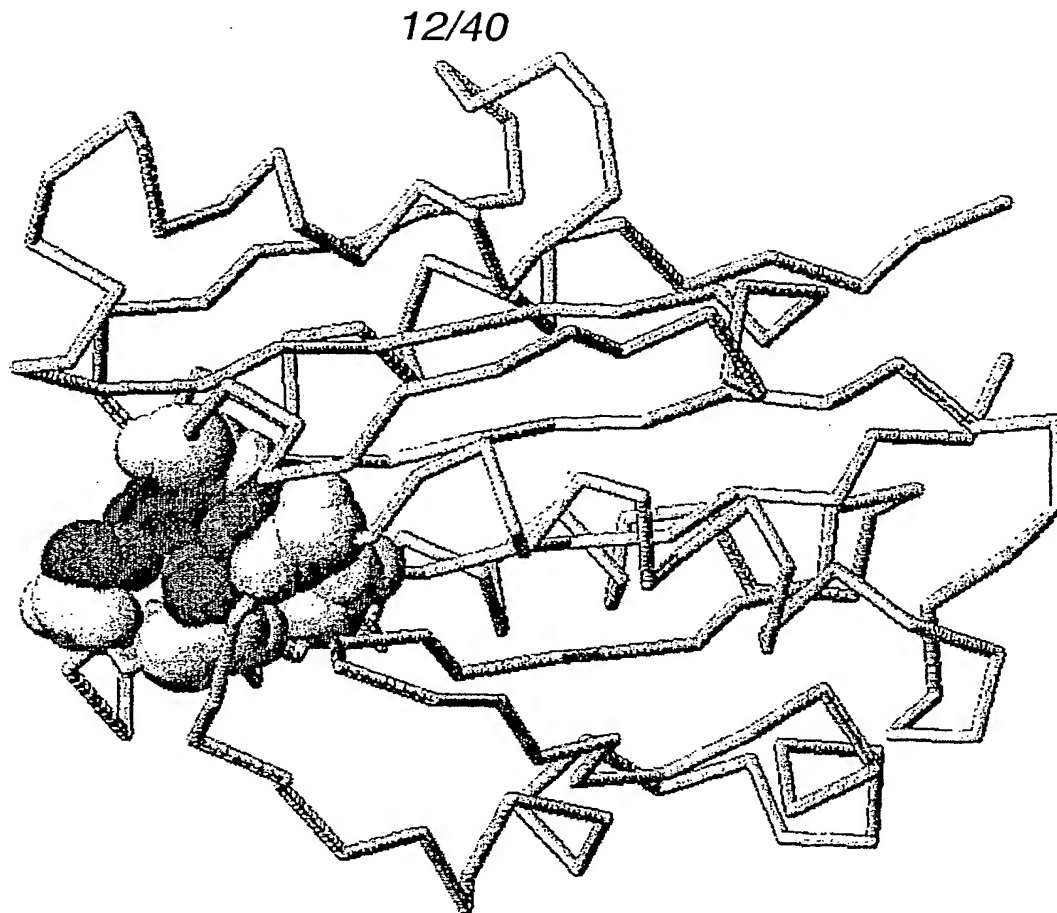
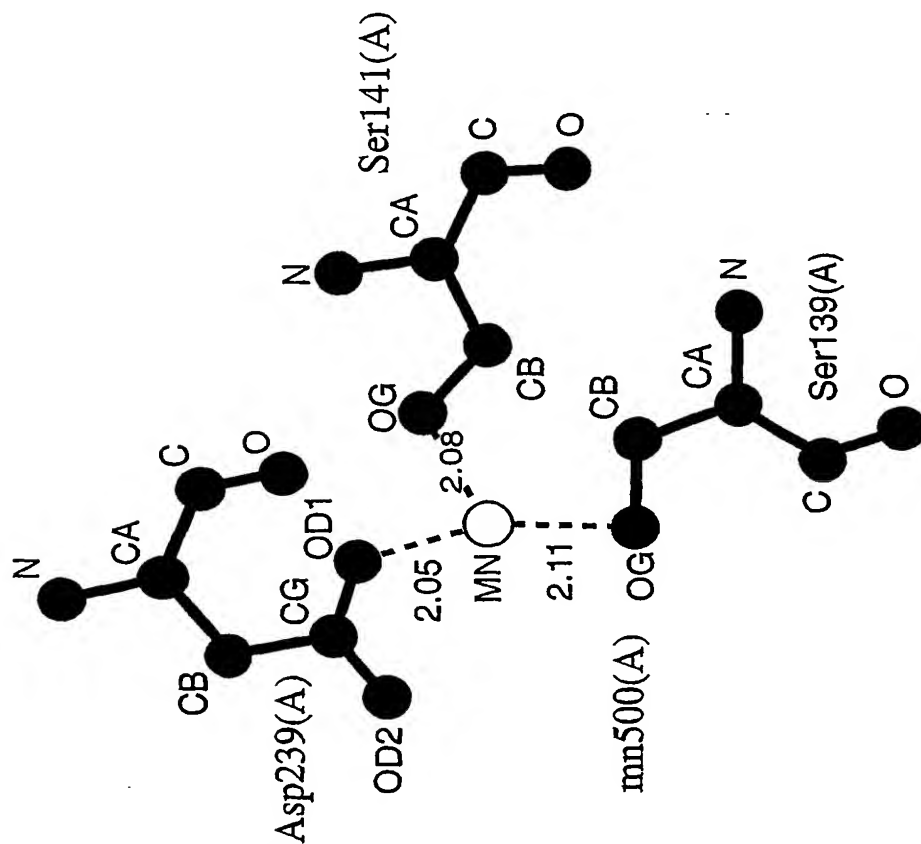


FIG. 8A

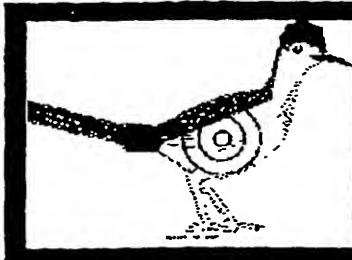


11fa: MN500 Chain [A]


FIG. 9

13/40

File Edit View Go Communicator Help  
Bookmarks Location http://victoria.inpharmatica.co.uk/~volker/BPD3target.html



Target Mining Interface



Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA):

OR

- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

Select Database

Release:

Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first\_PB\_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

14/40

FIG. 10A

Accession	DPD info	WVWEnt	Title	Organism	Div	WID (GI/FS)	Query ion (GI/FS)	Repeat ion (GI/FS)	App. score (GI)	Conf. (GI)	1st iter. (FS)	Best iter. (FS)	Best E-value (FS)
✓	P58199 cdt through Ions0811111 Ref. Seq. View	P58192	INTEGRIN ALPHA-1 (LAMININ AND COLLAGEN RECEPTOR) (VLA-1) (CD49A)	Homo sapiens (Human)	PRI	51.8%, 51% unmaskedSW	2-200, 2-200	138-207, 139-337	487	100% unmaskedGI	1	2	2E-73
✓	P12301 cdt through Ions0811111 Ref. Seq. View	P12301	PLATELET MEMBRANE GLYCOPROTEIN IA PRECURSOR (GPIA) (COLLAGEN RECEPTOR) (INTEGRIN ALPHA-2) (VLA-2 ALPHA CHAIN) (CD49B)	Homo sapiens (Human)	PRI	99.5%, 100% unmaskedSW	2-200, 2-200	169-367, 169-367	471	100% unmaskedGI	1	1	1E-111
✓	AA01258.1 cdt through Ions0811111 Ref. Seq. View	AA01258.1	Integrin alpha-11 subunit precursor	Homo sapiens	PRI	45.2%, 45% unmaskedSW	2-200, 2-200	159-355, 159-355	469	100% unmaskedGI	1	2	4E-68
✓	AA051918.2 cdt through Ions0811111 Ref. Seq. View	AA051918.2	Integrin alpha 11 subunit precursor	Homo sapiens	PRI	45.2%, 45% unmaskedSW	2-200, 2-200	159-355, 159-355	469	100% unmaskedGI	1	2	4E-68
✓	P39715 cdt through Ions0811111 Ref. Seq. View	Q92715	COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR	Homo sapiens (Human)	PRI	28.1%, 31% unmaskedSW	2-186, 7-189	133-318, 2323-2495	437	100% unmaskedGI	1	2	2E-42
✓	AA059181.1 cdt through Ions0811111 Ref. Seq. View	AA059181.1	Not given	Homo sapiens	PRI	27.1%, 27% unmaskedSW	2-200, 2-200	144-338, 144-338	435	100% unmaskedGI	1	2	1E-60
✓	AA059544.1 cdt through Ions0811111 Ref. Seq. View	AA059544.1	Not given	Homo sapiens	PRI	27.1%, 27% unmaskedSW	2-200, 2-200	144-338, 144-338	435	100% unmaskedGI	1	2	1E-60
✓	CA071722.1 cdt through Ions0811111 Ref. Seq. View	CA071722.1	α2(β3)D15.1 (collagen, type XII, alpha 1)	Homo sapiens	PRI	26.9%, 26% unmaskedSW	5-195, 156-201	293-472, 1581-1626	424	100% unmaskedGI	2	3	5E-6
✓	P21941 cdt through Ions0811111 Ref. Seq. View	P21941	CARTILAGE MATRIX PROTEIN PRECURSOR (MATRILIN-1)	Homo sapiens (Human)	PRI	26.8%, 26% unmaskedSW	3-189, 3-198	271-455, 37-224	424	100% unmaskedGI	1	2	3E-46
✓	AA011506.1 cdt through Ions0811111 Ref. Seq. View	AA011506.1	type XII collagen	Homo sapiens	PRI	27.6%, 30% unmaskedSW	2-185, 2-195	133-318, 133-318	422	100% unmaskedGI	1	2	3E-50
✓	AA031852.1 cdt through Ions0811111 Ref. Seq. View	AA031852.1	Integrin subunit alpha 10 precursor	Homo sapiens	PRI	56.5%, 46% unmaskedSW	2-199, 2-199	162-359, 162-359	419	100% unmaskedGI	1	2	5E-63
✓	CA003693.1 cdt through Ions0811111 Ref. Seq. View	CA003693.1	matrin-4	Homo sapiens	PRI	26%, 26% unmaskedSW	1-199, 2-200	335-529, 28-223	411	100% unmaskedGI	1	2	1E-47

2) 82 additional hits identified by both, Genome Threader and PSI-BLAST:

Crossed Genome Threader and PSI-BLAST values are shown in red.



15/40

FIG. 10B

GT  
Confidence

	Top50BlastHits Red. Seq. View	BAA91707.1	Not given	Homo sapiens	PRI	unmasked SW	unmasked SW	unmasked GT	Confidence	3	5	2E-8
J	AAA36795.1 drill through Top50BlastHits Red. Seq. View	AAA36795.1	undulin 2	Homo sapiens	PRI	34.7%, 21%	2-46, 139-196	100%	142	2	2	8E-6
J	AAD40367.1 drill through Top50BlastHits Red. Seq. View	AAD40367.1	calcium-activated chloride channel-2	Homo sapiens	PRI	15%, 15%	8-117, 8-117	100%	117	5	5	1E-4
J	CAA67559.1 drill through Top50BlastHits Red. Seq. View	CAA67559.1	collagen VI- $\alpha$ -1 chain	Homo sapiens	PRI	21.1%, 21%	108-197, 108-197	100%	114	3	3	5E-7
→	AAC76768.1 drill through Top50BlastHits Red. Seq. View	AAC76768.1	orf, hypothetical protein	Escherichia coli	BCT	13.2%, 11%	9-134, 7-186	100%	107	-4	-5	3E-24
J	CAB43000.1 drill through Top50BlastHits Red. Seq. View	CAB43000.1	tellurite resistance	Escherichia coli	BCT	12.9%, 14%	8-192, 2-181	100%	102	-6	-7	6E-22
J	AAA60114.1 drill through Top50BlastHits Red. Seq. View	AAA60114.1	platelet membrane glycoprotein IIb	Homo sapiens	PRI	14%, 13%	157-199, 157-199	98.52%	111	-7	-8	8E-5



17/40

**Inpharmatica Biopendium 1.4**  
File Edit Search Type Help

**Redundant Sequence Display**

Contains:  
2 Sub-sequences.  
0 PROSITE hits.  
0 PRINTS hits.

**Representative code: P03818 Length: 427 Organism name: Escherichia coli [additional annotation]**

Code	Aligned sequences	Code
P03818		P03818
AAC76768.1		AAC76768.1
AAA62097.1		AAA62097.1

**Primary database information**

<b>SWISS-PROT</b>	
Code	Details
P03818	Tax ID: 562
EC Number: Not available	Organism name: Escherichia coli
Gene Name: Y1EM	
Description: HYPOTHETICAL 49.6 KDA PROTEIN IN ASNA-KUP INTERGENIC REGION.	
Other links:	
ECOGENE: EG-H1730 PIR: A02443	

**GenBank**

protein_id	Details
AAC76768	Tax ID: 562
EC Number: Not available	Organism name: Escherichia coli
Gene Name: Y1EM	
Description: orf_hypothetical protein	

**Tools**

**Sequences**

**Ligands**

Start | BP04-F... | Inphar... | Fagen... | Inbox... | NCBI S... | Microso...

4:01 PM


FIG. 11

18/40

## FIG. 12


File Edit View Go Communicator Help

Bookmarks Location: <http://www.sanger.ac.uk/cgi-bin/Pfam/nph-search.cgi>



**Pfam**  
Protein families database of alignments and HMMs

Home | Keyword search | Protein search | DNA search | Browse Pfam | Taxonomy search | Help

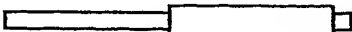


### Results for gi|2367274|gb|AAC76768.1|

There were no matches to Pfam-A (including borderline matches) for gi|2367274|gb|AAC76768.1|

Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B 15204	204	408	2.4e-108	Align


[427 residues]

---

### Alignments of Pfam-B domains to best-matching Pfam-B sequence

Format for fetching alignments to Pfam-B families: ☒ Hypertext linked to swisspfam ☐

Query gi|2367274|gb|AAC76768.1|204-408 matching Pfam-B 15204

```

YIEM_ECOLI 204 DILRLPPPELATLGITELEYFYRRLVEKQLLYRLNGESWREKVIERPV 253
                  DILRLPPPELATLGITELEYFYRRLVEKQLLYRLNGESWREKVIERPV
gi|2367274|gb|AAC76768.1| 204 DILRLPPPELATLGITELEYFYRRLVEKQLLYRLNGESWREKVIERPV 253

YIEM_ECOLI 254 VHKDYDEQPRGPFIVCVDTSGSHG6FHEQCAKAPCLALMRIALAEHRRCT 303
                  VHKDYDEQPRGPFIVCVDTSGSHG6FHEQCAKAPCLALMRIALAEHRRCT
gi|2367274|gb|AAC76768.1| 254 VHKDYDEQPRGPFIVCVDTSGSHG6FHEQCAKAPCLALMRIALAEHRRCT 303

YIEM_ECOLI 304 IMLFSTEIVRYELSGPQ6IEQAIRFLSQQFR6GTDLASCFRAIMERLQSR 353
                  IMLFSTEIVRYELSGPQ6IEQAIRFLSQQFR6GTDLASCFRAIMERLQSR
gi|2367274|gb|AAC76768.1| 304 IMLFSTEIVRYELSGPQ6IEQAIRFLSQQFR6GTDLASCFRAIMERLQSR 353

YIEM_ECOLI 354 EWFDAADVVISDFIAQRLPDDVTSKVKELQRVHQHMFHAVMSAMGKPGI 403
                  EWFDAADVVISDFIAQRLPDDVTSKVKELQRVHQHMFHAVMSAMGKPGI
gi|2367274|gb|AAC76768.1| 354 EWFDAADVVISDFIAQRLPDDVTSKVKELQRVHQHMFHAVMSAMGKPGI 403

YIEM_ECOLI 404 MRIFD 408
                  MRIFD
gi|2367274|gb|AAC76768.1| 404 MRIFD 408

```

Align to family

---

If you think there is anything wrong with this script, please contact [Pfam](#)

FIG. 13

19/40

File Edit View Go Communicator Help	
Bookmarks Location: <a href="http://www.ncbi.nlm.nih.gov/80/entrez/query.fcgi?cmd=Retrieve&amp;db=protein">http://www.ncbi.nlm.nih.gov/80/entrez/query.fcgi?cmd=Retrieve&amp;db=protein</a>	
LOCUS	ARC76768 427 aa BCT 01-DEC-2000
DEFINITION	orf, hypothetical protein [Escherichia coli K12].
ACCESSION	ARC76768
PIR	g2367274
VERSION	ARC76768.1 GI:2367274
DBSOURCE	locus AE000451 accession AE000451.1
KEYWORDS	
SOURCE	Escherichia coli K12.
ORGANISM	Escherichia coli K12 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE	1 (residues 1 to 427)
AUTHORS	Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, R.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Maphew, G.F., Gregor, J., Davis, R.W., Kirkpatrick, M.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.
TITLE	The complete genome sequence of Escherichia coli K-12
JOURNAL	Science 277 (5331): 1453-1474 (1997)
MEDLINE	97426617
PUBMED	9278503
REFERENCE	2 (residues 1 to 427)
AUTHORS	Blattner, F.R.
TITLE	Direct Submission
JOURNAL	Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE	3 (residues 1 to 427)
AUTHORS	Blattner, F.R.
TITLE	Direct Submission
JOURNAL	Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE	4 (residues 1 to 427)
AUTHORS	Plunkett, G. III.
TITLE	Direct Submission
JOURNAL	Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT	This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants M00301 and M001428 (from the Human Genome Project and HCHGR). The entire sequence was independently determined from E. coli K12 strain M61655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amherst.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with cc site nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.
FEATURES	Method: conceptual translation.
source	Location/Qualifiers 1..427 /organism="Escherichia coli K12" /strain="K12" /sub_strain="M61655" /db_xref="taxon:83333"
Protein	1..427 /function="orf; Unknown" /product="orf, hypothetical protein"
CDS	1..427 /gene="pieM" /coded_by="complement(2367272:5249..6532)" /transl_table=11 /note="f427; sequence change joins ORFs pieD and pieM from earlier version"
ORIGIN	1 mrsrlkdarv ppelteevmc yqqsqllstp qfiavqlpqil dlhlrlnsyw aeqarglvda 61 nstitsalht lflgrwrsls lvgattingq lleeereqll sevgermtls gqlepiladm 121 ntaagrlwda sagqlkrgdy qlivkygefi neqpelkrla eqglrsreak sipmdagae 181 tfrtmvrepv tpeqvvgllq qsdilrlly belatlcite leueferrly ehalltrrh

20/40

FIG. 14A

File Edit Search Type Help

# Accession Code Query

informatics

Tools

- ←
- 
- ?
- ?
- ?
- ?
- X

Sequences

- ? ?
- ? ?
- ? ?
- ? Acc y123

Ligands

- ? RDP NAD FAD
- ? Phe R1a Glu

Select a database and enter a code:

Database	
GenBank	
Accession	<input type="radio"/>
Accession Version	<input type="radio"/>
GI-ND	<input type="radio"/>
protein_id	<input type="radio"/>
GI-PID	<input type="radio"/>
SWISS-PROT	
Accession	<input type="radio"/>
ID	<input type="radio"/>
PDB	
ID	<input type="radio"/>

AAC76768.1

Submit Reset



FIG. 14B

File Edit Search Type Help

Aligned Sequence Display

Query details:  
AAC76768.1:orf, hypothetical protein

Tools

Sequences

Ligands

Page 1 of 3

Total hits: 88

Selected: 0

Total selected: 0

View alignment

Select all

Deselect all

Inpharmatica Genome Threader results

Cluster	Detail	Code	Title	Organism	%ID	Query rpn	Target rpn	Aln. score	Method	Confidence
1	1CK4B	CRYSTAL STRUCTURE OF RAT A1B1 INTEGRIN	CRYSTAL STRUCTURE OF RAT A1B1 INTEGRIN	Rhovegicus	12.6	267 - 398	6 - 144	158	Local	Certain(100%)
2	1IDO	1-DOMAIN FROM INTEGRIN CR3, MG2+ BO	1-DOMAIN FROM INTEGRIN CR3, MG2+ BO	Hsaplens	19.8	267 - 384	5 - 130	157	Local	Certain(100%)
3	1JLM	1-DOMAIN FROM INTEGRIN CR3, MN2+ BO	1-DOMAIN FROM INTEGRIN CR3, MN2+ BO	Hsaplens	19.8	267 - 384	5 - 130	156	Local	Certain(100%)
4	1CK4A	CRYSTAL STRUCTURE OF RAT A1B1 INTEGRIN	CRYSTAL STRUCTURE OF RAT A1B1 INTEGRIN	Rhovegicus	12.6	267 - 398	6 - 144	149	Local	Certain(100%)
5	1BHO1	MAC-1 DOMAIN MAGNESIUM COMPLEX	MAC-1 DOMAIN MAGNESIUM COMPLEX	Hsaplens	19.8	267 - 384	5 - 130	146	Local	Certain(100%)
6	1BHQ2	MAC-1 DOMAIN MAGNESIUM COMPLEX	MAC-1 DOMAIN MAGNESIUM COMPLEX	Hsaplens	19.8	267 - 384	5 - 130	146	Local	Certain(100%)
7	1BHQ1	MAC-1 DOMAIN MAGNESIUM COMPLEX	MAC-1 DOMAIN MAGNESIUM COMPLEX	Hsaplens	19.8	267 - 384	5 - 130	146	Local	Certain(100%)
8	1JDN2	MAC-1 DOMAIN METAL FREE	MAC-1 DOMAIN METAL FREE	Hsaplens	19.8	267 - 384	5 - 130	146	Local	Certain(100%)
9	1JDN1	MAC-1 DOMAIN METAL FREE	MAC-1 DOMAIN METAL FREE	Hsaplens	19.8	267 - 384	5 - 130	146	Local	Certain(100%)
10	1BHO2	MAC-1 DOMAIN MAGNESIUM COMPLEX	MAC-1 DOMAIN MAGNESIUM COMPLEX	Hsaplens	19.8	267 - 384	5 - 130	146	Local	Certain(100%)
11	1LFA8	CD11A 1-DOMAIN WITH BOUND MN++	CD11A 1-DOMAIN WITH BOUND MN++	Hsaplens	15.9	264 - 381	1 - 124	143	Local	Certain(100%)
12	1ZON	CD11A 1-DOMAIN WITHOUT BOUND CAT++	CD11A 1-DOMAIN WITHOUT BOUND CAT++	Hsaplens	15.9	264 - 381	1 - 124	143	Local	Certain(100%)
13	1LFAA	CD11A 1-DOMAIN WITH BOUND MN++	CD11A 1-DOMAIN WITH BOUND MN++	Hsaplens	14.6	264 - 354	1 - 96	141	Local	Certain(100%)
14	1ZOOA	CD11A 1-DOMAIN WITH BOUND MAGNESI	CD11A 1-DOMAIN WITH BOUND MAGNESI	Hsaplens	14.6	264 - 354	1 - 96	141	Local	Certain(100%)
15	1ZOPB	CD11A 1-DOMAIN WITH BOUND MAGNESI	CD11A 1-DOMAIN WITH BOUND MAGNESI	Hsaplens	14.6	264 - 354	1 - 96	141	Local	Certain(100%)
16	1ZOPA	CD11A 1-DOMAIN WITH BOUND MAGNESI	CD11A 1-DOMAIN WITH BOUND MAGNESI	Hsaplens	14.6	264 - 354	1 - 96	141	Local	Certain(100%)
17	1ZOOB	CD11A 1-DOMAIN WITH BOUND MAGNESI	CD11A 1-DOMAIN WITH BOUND MAGNESI	Hsaplens	14.6	264 - 354	1 - 96	141	Local	Certain(100%)
18	1DQA	NMR SOLUTION STRUCTURE OF THE INSER	NMR SOLUTION STRUCTURE OF THE INSER	Hsaplens	14.3	267 - 354	10 - 100	139	Local	Certain(100%)
19	1ATZA	HUMAN VON WILLEBRAND FACTOR A3 DO	HUMAN VON WILLEBRAND FACTOR A3 DO	Hsaplens	11.3	267 - 407	6 - 152	135	Local	Certain(100%)
20	1OAKA	CRYSTAL STRUCTURE OF THE VON WILLE	CRYSTAL STRUCTURE OF THE VON WILLE	Hsaplens	11.3	267 - 398	10 - 150	131	Local	Certain(100%)
21	1AOXA	1-DOMAIN FROM INTEGRIN ALPHA2-BETA1	1-DOMAIN FROM INTEGRIN ALPHA2-BETA1	N/A	13.9	267 - 398	5 - 143	107	Local	Certain(100%)
22	1GCSB	1-DOMAIN FROM INTEGRIN ALPHA1-BETA1	1-DOMAIN FROM INTEGRIN ALPHA1-BETA1	N/A	11.9	267 - 398	7 - 145	102	Local	Certain(100%)
23	1GCSA	1-DOMAIN FROM INTEGRIN ALPHA1-BETA1	1-DOMAIN FROM INTEGRIN ALPHA1-BETA1	N/A	11.9	267 - 398	7 - 145	102	Local	Certain(100%)
24	1AOXB	1-DOMAIN FROM INTEGRIN ALPHA2-BETA1	1-DOMAIN FROM INTEGRIN ALPHA2-BETA1	Hsaplens	10.7	267 - 398	8 - 145	98	Local	Certain(100%)
25	1A03A	A3 DOMAIN OF VON WILLEBRAND FACTOR	A3 DOMAIN OF VON WILLEBRAND FACTOR	Hsaplens	10.7	267 - 398	8 - 145	98	Local	Certain(100%)
26	1A03B	A3 DOMAIN OF VON WILLEBRAND FACTOR	A3 DOMAIN OF VON WILLEBRAND FACTOR	Hsaplens	10.7	267 - 398	8 - 145	98	Local	Certain(100%)
27	1AUQ	A1 DOMAIN OF VON WILLEBRAND FACTOR	A1 DOMAIN OF VON WILLEBRAND FACTOR	Hsaplens	11.8	250 - 398	4 - 158	97	Local	Certain(100%)
28	1ATZB	HUMAN VON WILLEBRAND FACTOR A3 DO	HUMAN VON WILLEBRAND FACTOR A3 DO	Hsaplens	10.7	267 - 398	8 - 146	96	Local	Certain(100%)
29	1A6JB	NITROGEN REGULATORY BACTERIAL PRO	NITROGEN REGULATORY BACTERIAL PRO	Ecol	18.6	306 - 349	13 - 57	60	Local	Low(42%)
30	1BKDS	COMPLEX OF HUMAN H-RAS WITH HUMA	COMPLEX OF HUMAN H-RAS WITH HUMA	Hsaplens	14.4	141 - 292	100 - 202	59	Local	Low(34%)

GenBank

SWISS-PROT

PDB

File Edit Search Type Help

# Aligned Sequence Display

Query details:

AAC76768.1: orf, hypothetical protein



Tools



View alignment

Total selected: 0

Reverse Maximised PSI-BLAST results:

Select all

Deselect all

Total hits: 508

Selected: 0

Page 1 of 17

Cluster	Detail	Code	Title	Organism	%ID	Query rgn	Target rgn	Filter	E value
0	BAA80016.1	452aa long hypothetical protein		<i>A. permix</i>	28.0	156 - 407	202 - 431	2	2.00E-50
0	Q80364	HYPOTHETICAL PROTEIN MJ0077		<i>M. jamaicensis</i>	22.0	146 - 408	122 - 361	3	2.00E-71
0	AAB88810.1	conserved hypothetical protein		<i>M. jamaicensis</i>	23.0	147 - 410	170 - 420	3	2.00E-65
0	AAG78910.1	putative retroelement polyploid protein		<i>A. thaliana</i>	18.0	168 - 404	144 - 403	3	2.00E-50
0	Q55874	HYPOTHETICAL 458 KDA PROTEIN SL0103		<i>S. sp.</i>	21.0	241 - 407	17 - 168	3	4.00E-38
0	AAB84857.1	magnesium chelatase subunit Chl		<i>M. thermotrophicum</i>	25.0	151 - 370	324 - 535	3	5.00E-22
0	CAA90478.1	Not given		<i>M. musculus</i>	20.0	247 - 416	132 - 309	4	2.00E-44
0	AA668544.1	Not given		<i>H. sapiens</i>	16.0	223 - 412	107 - 305	4	3.00E-42
0	AAB24821.1	leukocyte integrin alpha chain		<i>H. sapiens</i>	16.0	223 - 412	107 - 305	4	3.00E-42
0	1JEM	I-DOMAIN FROM INTEGRIN CR3, MN2+ BOUND		<i>H. sapiens</i>	16.0	263 - 412	1 - 158	4	2.00E-41
0	1IDO	I-DOMAIN FROM INTEGRIN CR3, MG2+ BOUND		<i>H. sapiens</i>	16.0	263 - 412	1 - 158	4	2.00E-41
0	1BHO.1	MAG-1 DOMAIN MAGNESIUM COMPLEX		<i>H. sapiens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
0	1IDN2	MAG-1 DOMAIN METAL FREE		<i>H. sapiens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
0	1IDN1	MAG-1 DOMAIN METAL FREE		<i>H. sapiens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
0	1BHO.2	MAG-1 DOMAIN MAGNESIUM COMPLEX		<i>H. sapiens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
0	1BHO.3	MAG-1 DOMAIN MAGNESIUM COMPLEX		<i>H. sapiens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
0	1BHO.1	MAG-1 DOMAIN MAGNESIUM COMPLEX		<i>H. sapiens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
0	Q22437	MAGNESIUM-CHLATASE SUBUNIT CHL2 PRECURSOR		<i>P. salivum</i>	18.0	162 - 417	458 - 701	4	7.00E-41
0	CAB68268.1	Not given		unidentified	18.0	162 - 417	473 - 705	4	2.00E-40
0	AAB16889.1	CG11b		<i>S. sp.</i>	17.0	254 - 417	5 - 177	4	8.00E-40
0	CAB68178.1	Mg-protoporphylin IX		<i>H. vulgare</i>	17.0	162 - 410	90 - 311	4	2.00E-38
0	AAD52031.1	magnesium-chelatase subunit CHL2 precursor		<i>A. thaliana</i>	17.0	162 - 423	441 - 695	4	5.00E-38
0	BAA16787.1	Mg chelatase subunit		<i>S. sp.</i>	16.0	189 - 367	415 - 586	4	6.00E-38
0	AAB85281.1	conserved protein		<i>M. thermotrophicum</i>	13.0	154 - 393	50 - 303	4	7.00E-37
0	CAA77537.1	561 aa (60 kD) Mg chelatase subunit		<i>R. capsulatus</i>	21.0	168 - 367	277 - 484	4	1.00E-36
0	AAB22885.1	T27G7.20		<i>A. thaliana</i>	17.0	162 - 423	491 - 747	4	2.00E-36
0	AAF21241.1	alpha D integrin		<i>R. norvegicus</i>	14.0	238 - 412	128 - 307	4	2.00E-35
0	AAF23482.1	leukocyte adhesion glycoprotein p150.85 alpha integrin subunit		<i>M. musculus</i>	13.0	231 - 412	118 - 307	4	2.00E-35
0	AAB36547.1	leukointegrin alpha d chain		<i>H. sapiens</i>	13.0	238 - 410	124 - 303	4	3.00E-35

GenBank SWISS-PROT PDB

23/40

## FIG. 15

AlEye output (December 13, 2000 3:07 PM)

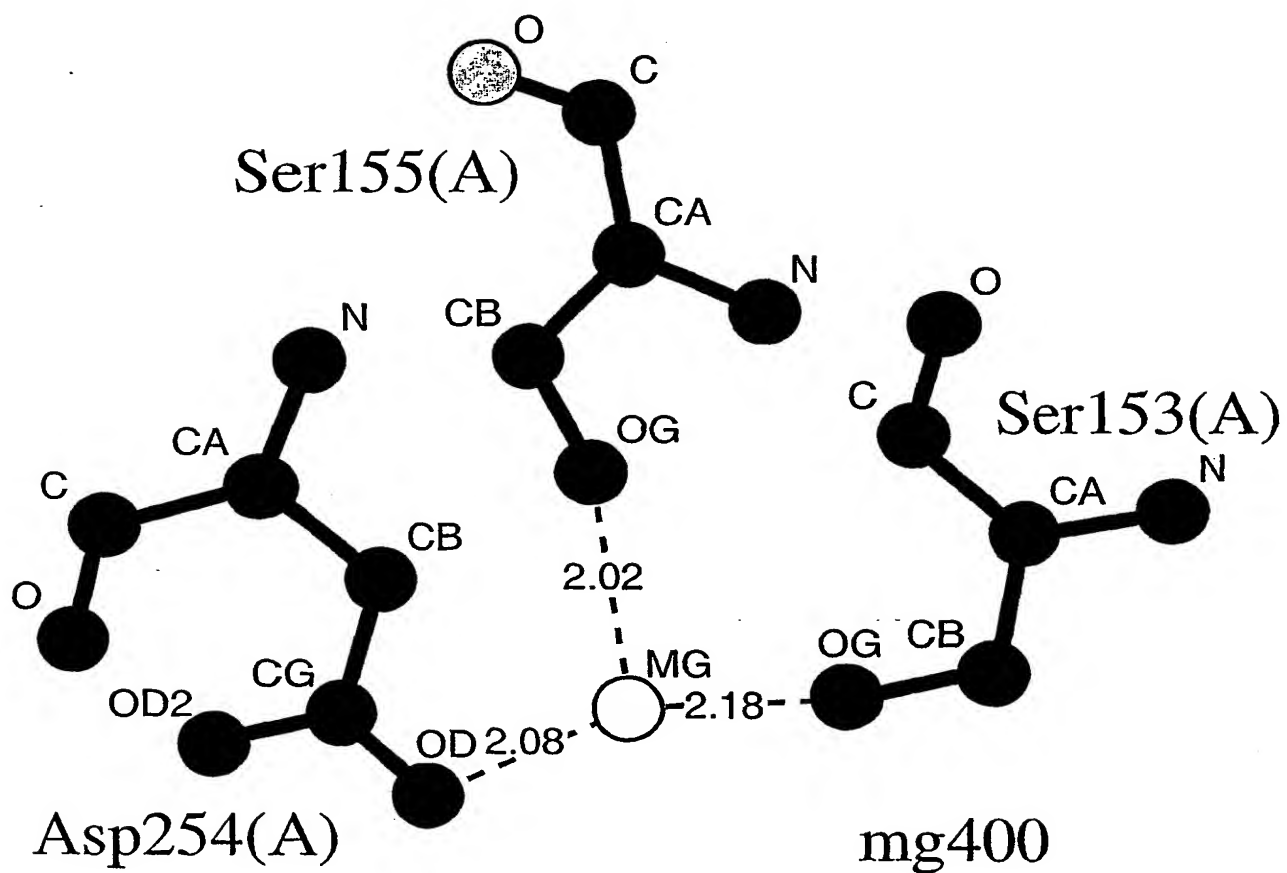
```

AAC76768.1 10 20 30 40 50 60 70
1A0X:A MRSRLKDA RVPPELT EEVN C Y Q S Q L L S T P Q F I V Q L P Q I L D L L H R L N S P W A E Q A R Q L V D A N S T I T S A L H T L F L Q R W R
-----
AAC76768.1 80 90 100 110 120 130 140 150
1A0X:A L S I V Q A T T L N Q Q L L E E E R E Q L L S E V Q E R H T L S G Q L E P I L A D N N T A A G R L W D M S A G Q L K R G D Y Q L I V K Y G E F L N E Q P
-----
AAC76768.1 160 170 180 190 200 210 220 230
1A0X:A E L K R L A E Q L G R S R E A K S I P P N D A Q M E T F R T M V R E P A T V P E Q V D G L Q Q S D D I L R L L P P E L A T L G I T E L E Y E F Y P R L V E
-----
AAC76768.1 240 250 260 270 280 290 300
1A0X:A K Q L L T Y R L H G E S W R E K V I E R P V V H K D Y D E Q P R G P F I V C V D T S G S M G G F N E Q C A K A F C L A L M R I - - A L A E N R R C Y I M L
-----
AAC76768.1 310 320 330 340 350 360 370 380
1A0X:A - - - - - S C P S L I D - - - - - V V Y V C D E S N S I Y P W D A - - V K N F L E K F V Q G L D I G P T K T Q V G L I Q
-----
AAC76768.1 390 400 410 420 430 440 450 460 470
1A0X:A F S T E I - V R Y E L S G P Q G - - - I E Q A I R F L S Q Q F R G G T O L A S C F R A I M E R L Q S - - - - - R E W F D A D A V V I S D F I A Q R L P D D
Y A N N P R V V F N L N T Y K T K E E M I V A T S Q T S Q Y G G D L N T F G A I Q Y A R K Y A Y S A A S G G R R S A T K Y M V V V T D G E S H - D G S M
-----
AAC76768.1 480
1A0X:A V T S K V K E L Q R V H Q H R F H A V A M S A H G K P G I M R I F D H I W R F D T G M R S R L L R W R R
L K A V I D Q C N H D N - I L R F G I A V L G Y - - - - - L N R N A L D T K N L I K E I K A I A S I P T E R Y F F N V S D E A A L L E K A G
-----
AAC76768.1
1A0X:A T L G E Q I F S I E G G T

```

24/40

FIG. 16A



1aox: MG400

25/40

FIG. 16B

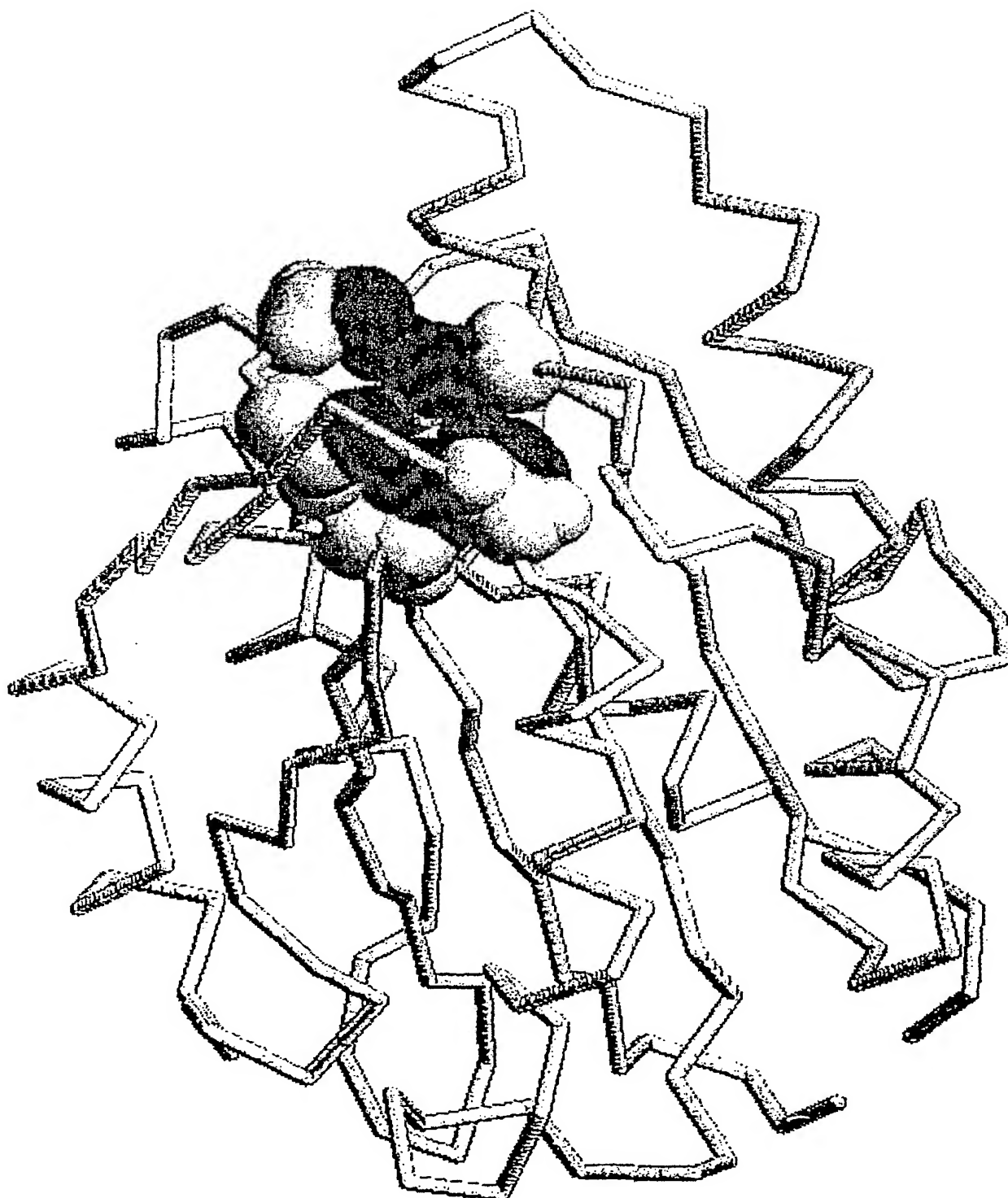
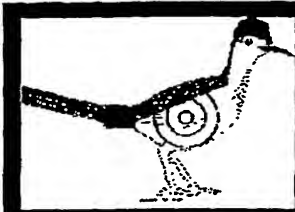



FIG. 17

26/40

File Edit View Go Communicator Help  
Bookmarks Location: http://victoria.inpharmatica.co.uk/~volker/BPD3target.html

**Target Mining Interface**



### Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA):  and chain (e.g. B):
- OR
- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

### Select Database

Release:

### Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first\_PB\_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

- Filter for the following SPECIES:  
☐ Homo sapiens ☐ Rattus norvegicus (Rat) ☐ Mus musculus (Mouse) ☐ Danio rerio (Zebra fish)

100%

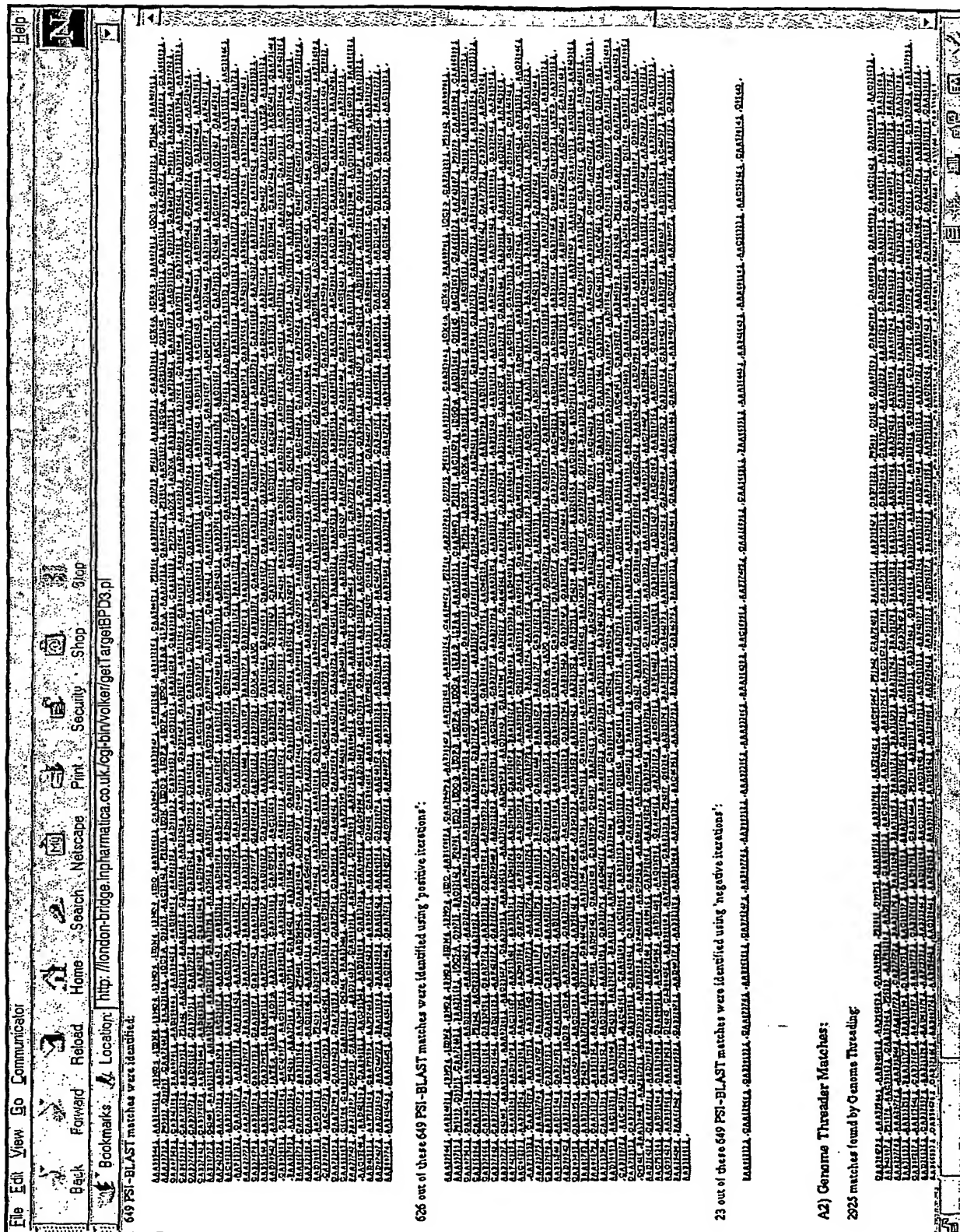


FIG. 18A

27/40

File Edit View Go Communicator Help											
Back Forward Reload Home Search Netscape Print Security Stop											
Bookmarks Location: http://London-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl											
2) 81 additional hits identified by both, Genome Threader and PSI-BLAST:											
Combined Genome Threader and PSI-Blast output: PSI-BLAST values are shown in maroon!											
Add2list	BPD links	WWW link	Title	Organism	Div.	%ID (GT,PSI)	Query rgn. (GT,PSI)	Target rgn. (GT,PSI)	Aln. score (GT)	Conf. (GT)	1st Iter. (PSI)
J	<a href="#">AAA59544.1</a> <a href="#">drill through</a> <a href="#">Top50BlastHits</a> <a href="#">Red Seq View</a>	<a href="#">AAA59544.1</a>	Not given	Homo sapiens	PRI	100%, 100% unmaskedSW	1-187, 1-187	148-334, 148-334	488	100% unmaskedGT	1
→	<a href="#">AAB24821.1</a> <a href="#">drill through</a> <a href="#">Top50BlastHits</a> <a href="#">Red Seq View</a>	<a href="#">AAB24821.1</a>	leukocyte integrin alpha chain	Homo sapiens	PRI	100%, 100% unmaskedSW	1-187, 1-187	148-334, 148-334	488	100% unmaskedGT	1
J	<a href="#">Q99715</a> <a href="#">drill through</a> <a href="#">Top50BlastHits</a> <a href="#">Red Seq View</a>	<a href="#">Q99715</a>	COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR.	Homo sapiens (Human)	PRI	28.9%, 26% unmaskedSW	2-186, 2-179	439-617, 2322-2494	456	100% unmaskedGT	1
J	<a href="#">AAB38702.1</a> <a href="#">drill through</a> <a href="#">Top50BlastHits</a> <a href="#">Red Seq View</a>	<a href="#">AAB38702.1</a>	cartilage matrix protein	Homo sapiens	PRI	28.9%, 25% unmaskedSW	2-186, 2-186	274-452, 40-221	446	100% unmaskedGT	1
J	<a href="#">AAC01506.1</a> <a href="#">drill through</a> <a href="#">Top50BlastHits</a> <a href="#">Red Seq View</a>	<a href="#">AAC01506.1</a>	type XII collagen	Homo sapiens	PRI	28.4%, 28% unmaskedSW	2-186, 2-186	137-318, 137-318	445	100% unmaskedGT	1
J	<a href="#">CAA72402.1</a> <a href="#">drill through</a> <a href="#">Top50BlastHits</a> <a href="#">Red Seq View</a>	<a href="#">CAA72402.1</a>	collagen type XIV	Homo sapiens	PRI	28.7%, 30% unmaskedSW	2-186, 2-186	6-185, 6-185	442	100% unmaskedGT	1
J	<a href="#">AAB38547.1</a> <a href="#">drill through</a> <a href="#">Top50BlastHits</a> <a href="#">Red Seq View</a>	<a href="#">AAB38547.1</a>	leukocyte integrin alpha d chain	Homo sapiens	PRI	61%, 60% unmaskedSW	1-187, 1-187	148-334, 148-334	439	100% unmaskedGT	1
J	<a href="#">CAB71222.1</a> <a href="#">drill through</a> <a href="#">Top50BlastHits</a> <a href="#">Red Seq View</a>	<a href="#">CAB71222.1</a>	dJ238D15.1 (collagen, type XII, alpha 1)	Homo sapiens	PRI	27.1%, 22% unmaskedSW	1-186, 2-186	293-472, 1430-1620	439	100% unmaskedGT	1
J	<a href="#">CAA07569.1</a> <a href="#">drill through</a> <a href="#">Top50BlastHits</a> <a href="#">Red Seq View</a>	<a href="#">CAA07569.1</a>	matrin-4	Homo sapiens	PRI	27.9%, 24% unmaskedSW	2-186, 3-187	344-525, 34-215	418	100% unmaskedGT	2
J	<a href="#">CAB46380.1</a> <a href="#">drill through</a> <a href="#">Top50BlastHits</a> <a href="#">Red Seq View</a>	<a href="#">CAB46380.1</a>	dJ453C12.3 (matrin-4)	Homo sapiens	PRI	27.9%, 24% unmaskedSW	2-186, 3-186	385-566, 34-214	418	100% unmaskedGT	2





30/40

FIG. 19

Biopendium additional annotation page - Netscape

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Bookmarks Location: /london-bridge/inpharmatica/BPDEV1/cgi-bin/isp.pl?rep\_esn=55817&password=caiss\_app&username=caiss\_app&role\_sid=BPDEV1 What's Related

Aligned annotation view for P10155 (downloading image...)

1: A4435403.1  
Rep: P10155

1 50 100 150 200 250 300 350 400 450 500

Primary database information: ☒ **Prosite-matches** ☒ Prints matches

Secondary database information: ☒ **Phosphorylation** ☒ **Transmembrane regions**

Inpharmatica calculated information: ☒ **Transmembrane regions**

**Sequence information**

Source databank	SWISSPROT
Accession code	P10155
Gene name	SSA2 OR R060
Download sequence in FASTA format	

Links to other resources:

View custom hyperlinks

Search in ExPASy

European Bioinformatics Institute	PR000504
Online Mendelian Inheritance in Man	234700
Online Mendelian Inheritance in Man	60063
Protein Information Resource	A30596
Protein Information Resource	A31760

Swissprot comments:

FUNCTION: UNKNOWN

PRO SMALL RIBONUCLEOPROTEINS CONSIST OF FOUR SMALL RNA MOLECULES OF 85-112 NT EACH OF


Document Done

31/40  
FIG. 20

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Stop


Bookmarks Location: http://www.sanger.ac.uk/cgi-bin/Pfam/nph-search.cgi



# Pfam

Protein families database of alignments and HMMs

Home | Keyword search | Protein search | DNA search | Browse Pfam | Taxonomy search | Help



## Results for gi|133251|sp|P10155|RO60\_HUMAN

There were no matches to Pfam-A (including borderline matches) for gi|133251|sp|P10155|RO60\_HUMAN

### Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B_8344	1	194	2.3e-103	<a href="#">Align</a>
Pfam-B_10162	195	538	1.8e-165	<a href="#">Align</a>

[538 residues]

---

### Alignments of Pfam-B domains to best-matching Pfam-B sequence

Format for fetching alignments to Pfam-B families: ☐ [Hypertext linked to swisspfam](#)

Query gi|133251|sp|P10155|RO60\_HUMAN/1-194 matching Pfam-B\_8344

```

092787 1 MEESVNQMOPLEKQIANSDGYYVWQVTDNRLHRFLCFGSEGGTTYIKE 50
gi|133251|sp|P10155|RO60_HUMAN 1 MEESVNQMOPLEKQIANSDGYYVWQVTDNRLHRFLCFGSEGGTTYIKE 50
092787 51 QKLGLENAEALIRLIEDGRGCEVIQEIKSFSGEGRTTKQEPMLFALAI CS 100
gi|133251|sp|P10155|RO60_HUMAN 51 QKLGLENAEALIRLIEDGRGCEVIQEIKSFSGEGRTTKQEPMLFALAI CS 100
092787 101 QCSDISTKQARFKAVSEVCRIPTHLPTFIQFKKDLKESMKCGMWERLRK 150
gi|133251|sp|P10155|RO60_HUMAN 101 QCSDISTKQARFKAVSEVCRIPTHLPTFIQFKKDLKESMKCGMWERLRK 150
092787 151 AIADWYNEKGGMALALAVTKYKQRNGWSHKDLLRLSHLKPSSSG 194
gi|133251|sp|P10155|RO60_HUMAN 151 AIADWYNEKGGMALALAVTKYKQRNGWSHKDLLRLSHLKPSSSG 194
  
```

[Align to family](#)

Query gi|133251|sp|P10155|RO60\_HUMAN/195-538 matching Pfam-B\_10162

```

008848 195 LAIVTKYITKGWKEVMEETKEKALSVEREKLKYLEAVEKVKRTKDDLEV 244
gi|133251|sp|P10155|RO60_HUMAN 195 LAIVTKYITKGWKEVMEETKEKALSVEREKLKYLEAVEKVKRTKDDLEV 244
008848 245 IMLEENQGLVREHLLTNHLKSKEVWKALLQEMPLTALLRHLGKMTANSVL 294
THI.TERN+LVREHLLTNHLKSKEVWKALLQEMPLTALLRHLGKMTANSVL
008848 295 EPGHSEVSLICEKL SHEKLLKARIHPFHVLI ALETYTRAGHLRGLKWKY 344
gi|133251|sp|P10155|RO60_HUMAN 295 EPGHSEVSLICEKL SHEKLLKARIHPFHVLI ALETYTRAGHLRGLKWKY 344
008848 345 PDKDILQALDRAFYTTFTVTEPT6KRFLAVDVSAASNQRALGSVLHAST 394
  
```



32/40  
FIG. 21

File Edit View Go Communicator		Help
Back Forward Reload Home Search Netscape Print Security Stop		
Bookmarks Location: <a href="http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&amp;db=Prot">http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&amp;db=Prot</a>		
LOCUS	R060_HUMAN 538 aa PRI 01-FEB-1996	
DEFINITION	60 KD RO PROTEIN (60 KD RIBONUCLEOPROTEIN RO) (RORNP) (SJOGEN SYNDROME TYPE A ANTIGEN (SS-A)).	
ACCESSION	P10155	
PID	g133251	
VERSION	P10155 GI:133251	
DBSOURCE	swissprot: locus R060_HUMAN, accession P10155. class: standard. created: Mar 1, 1989. sequence updated: Mar 1, 1989. annotation updated: Feb 1, 1996. xrefs: gi: gi: 177782, gi: gi: 177783, gi: gi: 387656, gi: gi: 387657, gi: gi: 86722, gi: gi: 107626 xrefs (non-sequence databases): MIM 600053, MIM 234700, PROSITE P500030	
KEYWORDS	Ribonucleoprotein; RNA-binding; Systemic lupus erythematosus; Antigen.	
SOURCE	human.	
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (residues 1 to 538)	
AUTHORS	Deutscher, S.L., Marley, J.B. and Keene, J.D.	
TITLE	Molecular analysis of the 60-kDa human Ro ribonucleoprotein	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 85 (24), 9479-9483 (1988)	
MEDLINE	89071722	
REMARK	SEQUENCE FROM N.A.	
REFERENCE	2 (residues 1 to 538)	
AUTHORS	Ben-Chetrit, E., Gandy, B.J., Tan, E.M. and Sullivan, K.F.	
TITLE	Isolation and characterization of a cDNA clone encoding the 60-kD component of the human SS-A/Ro ribonucleoprotein autoantigen	
JOURNAL	J. Clin. Invest. 83 (4), 1284-1292 (1989)	
MEDLINE	89198084	
REMARK	SEQUENCE FROM N.A.	
COMMENT	<p>This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. The original entry is available from <a href="http://www.expasy.ch/sprot">http://www.expasy.ch/sprot</a> and <a href="http://www.ebi.ac.uk/sprot">http://www.ebi.ac.uk/sprot</a></p> <p>[FUNCTION] UNKNOWN.</p> <p>[SUBUNIT] RO SMALL RIBONUCLEOPROTEINS CONSIST OF FOUR SMALL RNA MOLECULES OF 85-112 NT, EACH OF WHICH IS COMPLEXED WITH A 60 KD PROTEIN. RO RNPS MAY ALSO CONTAIN AN ADDITIONAL 52 KD PROTEIN.</p> <p>[SUBCELLULAR LOCATION] CYTOPLASMIC.</p> <p>[DISEASE] SERA FROM PATIENTS WITH SYSTEMIC LUPUS ERYTHEMATOSUS OFTEN CONTAIN ANTIBODIES THAT REACT WITH THE NORMAL CELLULAR RO PROTEIN AS IF THESE ANTIGEN WAS FOREIGN.</p> <p>[SIMILARITY] CONTAINS 1 RNA RECOGNITION MOTIF (RNP).</p> <p>[SIMILARITY] STRONG, TO XENOPUS 60 KD RO PROTEIN.</p>	
FEATURES	Location/Qualifiers source 1..538 /organism="Homo sapiens" /db_xref="taxon:9606" Protein 1..538 /product="60 KD RO PROTEIN" Region 93..98 /region_name="Domain" /note="RNA-BINDING (RNP2) (BY SIMILARITY)." Region 124..131 /region_name="Domain" /note="RNA-BINDING (RNP1) (BY SIMILARITY)." Region 239 /region_name="Conflict" /note="K -> R (IN REF. 2)." Region 515..538 /region_name="Conflict" /note="GMLDMCGFDTCALDVIRNFTLDMI -> ALQNTLLHKSF (IN REF. 2)." 	
ORIGIN	1 meesvnmqgp lnekqiansq dgyvwqvtdm nrllhrflcsg seggtypike qklglенааа 61 lirliedgrg cevigeiksf sgegrttkqe pmlfalaics qosdistkqa afkavsever 121 ipthlftfiq fkkdlkesmk egmwgralrk aiadwynekg gmalalavtk pzqngwshk 181 dlrlrlshlkp sseglaivtk yitkgwkevh elykekalv etekllkyle avekvkrtdk 241 elevihliee hrlvrehllt nhlskskewk allqempita llrnlgkmta nsvlepgmse 301 vslvceklen ekllkkaarh pphilialet yktghglrgk lkwrpdeeil kaldaafpht 361 fktveptgkr fllavdvss mnqrvlgsil nastvaaame mvvtrtekds yvvaifsdemv 421 pcpvtttdntl qqvlmamsqi paggtdeclp miwaqktntp advfivftdn etfaggvghpa 481 ialreyrkkm dipaklivcg mtsngftiad pddrgmldmc gfdtgaldvi rnftldmi	



33/40

FIG. 22A

File Edit Search Type Help

Accession Code Query

informatics

Tools

Sequences

Ligands

Select a database and enter a code:

Database	
GenBank	
Accession	<input type="radio"/>
Accession Version	<input type="radio"/>
GI-NUD	<input type="radio"/>
protein Id	<input type="radio"/>
GI-PID	<input type="radio"/>
SWISS-PROT	
Accession	<input type="radio"/>
ID	<input type="radio"/>
PDB	
ID	<input type="radio"/>

P10155

Submit Reset

34/40

File Edit Search Type Help

FIG. 22B  
Aligned Sequence Display

Query details:

P10155: 60 KDA RO PROTEIN (60 KDA RIBONUCLEOPROTEIN RO) (RORNP) (SJORGEN SYNDROME TYPE A ANTIGEN (SS-A)).



Tools



Total selected: 1. View alignment

Inpharmatica Genome Threader results:

Page 1 of 2.

Total hits: 31

Selected: 1

Select all

Deselect all

Cluster Detail	Code	Title	Organism	%ID	Query len	Target len	Aln. score	Method	Confidence
1A5A:A	CRYO-CRYSTALLOGRAPHY OF A TRUE SU...	<i>Syphilium</i>	10.9	42	122	120-250	54	Local	Low(76%)
1JLM	I-DOMAIN FROM INTEGRIN CR3, MN2+ BO...	<i>Hsapilens</i>	11.1	373	504	6-145	55	Local	Low(70%)
1BHO1	MAC-1 I DOMAIN MAGNESIUM COMPLEX	<i>Hsapilens</i>	11.7	373	504	6-145	55	Local	Low(67%)
1BHO2	MAC-1 I DOMAIN MAGNESIUM COMPLEX	<i>Hsapilens</i>	11.7	373	504	6-145	55	Local	Low(67%)
1IDN2	MAC-1 I DOMAIN METAL FREE	<i>Hsapilens</i>	11.7	373	504	6-145	55	Local	Low(67%)
1BHQ2	MAC-1 I DOMAIN CADMIUM COMPLEX	<i>Hsapilens</i>	11.7	373	504	6-145	55	Local	Low(67%)
1BHQ1	MAC-1 I DOMAIN CADMIUM COMPLEX	<i>Hsapilens</i>	11.7	373	504	6-145	55	Local	Low(67%)
1IDN1	MAC-1 I DOMAIN METAL FREE	<i>Hsapilens</i>	11.7	373	504	6-145	55	Local	Low(67%)
1BEOA	TRP SYNTHASE (DBON-IPP-SER) WITH K+	<i>Syphilium</i>	10.3	24	172	102-248	53	Local	Low(58%)
1IDO	I-DOMAIN FROM INTEGRIN CR3, MG2+ BO...	<i>Hsapilens</i>	10.9	373	488	6-131	58	Local	Low(55%)
2TYS	CRYSTAL STRUCTURE OF MUTANT (BETA...	<i>Hsapilens</i>	11.5	25	174	103-252	56	Local	Low(53%)
TAOXA	I-DOMAIN FROM INTEGRIN ALPHA2-BETA1	<i>Hsapilens</i>	15.0	411	504	45-148	53	Local	Low(53%)
1UBSA	TRYPTOPHAN SYNTHASE (EC.4.2.1.20) WT...	<i>Syphilium</i>	11.5	42	174	120-254	65	Local	Low(48%)
1BKS	TRYPTOPHAN SYNTHASE (EC.4.2.1.20) FR...	<i>Syphilium</i>	10.9	42	172	120-251	61	Local	Low(36%)
1OAK	CRYSTAL STRUCTURE OF THE VON WILLE...	<i>Hsapilens</i>	6.7	366	504	4-150	49	Local	Low(34%)
1AOXB	I-DOMAIN FROM INTEGRIN ALPHA2-BETA1	<i>Hsapilens</i>	10.3	411	504	47-148	48	Local	Low(31%)
1GGC1	STRUCTURE OF THE COMPLEX OF AN FAB...	<i>Falisse</i>	21.1	277	355	10-100	62	Local	Low(23%)
1FMD1	FOOT AND MOUTH DISEASE VIRUS TYPE...	<i>Falisse</i>	21.1	277	355	10-100	62	Local	Low(22%)
1ALQ	A1 DOMAIN OF VON WILLEBRAND FACTOR	<i>Hsapilens</i>	7.0	373	504	20-159	43	Local	Low(21%)
1BWGB	CRYSTAL STRUCTURE OF TYROSINE AMIN...	<i>Taral</i>	5.5	463	536	178-248	57	Local	Low(15%)
1TTPA	TRYPTOPHAN SYNTHASE (EC.4.2.1.20) IN T...	<i>Syphilium</i>	10.4	44	172	122-251	51	Local	Low(15%)
1TTGA	TRYPTOPHAN SYNTHASE (EC.4.2.1.20) IN T...	<i>Syphilium</i>	10.4	44	172	122-251	51	Local	Low(15%)
1ART	ASPARTATE AMINOTRANSFERASE (EC.2.6...	<i>Ecol</i>	10.4	405	537	116-248	52	Local	Low(15%)
1GITA	ASPARTATE AMINOTRANSFERASE FROM...	<i>N/A</i>	10.5	405	537	116-248	45	Local	Low(12%)
1EEGA	CRYSTAL STRUCTURE OF YEAST KARYOP...	<i>Sacchari</i>	15.5	35	130	238-336	49	Local	Low(10%)
1EEGB	CRYSTAL STRUCTURE OF YEAST KARYOP...	<i>Sacchari</i>	15.5	35	130	238-336	49	Local	Low(10%)
1ARHA	ASPARTATE AMINOTRANSFERASE, Y225R/...	<i>Ecol</i>	10.5	405	537	116-248	46	Local	Low(10%)
1ARHB	ASPARTATE AMINOTRANSFERASE, Y225R/...	<i>Ecol</i>	10.5	405	537	116-248	46	Local	Low(10%)
1EFTA	CRYSTAL STRUCTURE OF THE MOESIN FE...	<i>N/A</i>	18.8	82	171	121-188	55	Local	Low(10%)
1TEF1B	CRYSTAL STRUCTURE OF THE MOESIN FE...	<i>N/A</i>	18.8	82	171	121-188	55	Local	Low(10%)

GenBank SWISS-PROT PDB

File Edit Search Type Help

# FIG. 22B

## Aligned Sequence Display

Query details:

P10155: 60 KDA RO PROTEIN (60 KDA RIBONUCLEOPROTEIN RO) (RORNP) (SJORGEN SYNDROME TYPE A ANTIGEN (SS-A)).



Tools:

Sequences:

Ligands:

Total selected: 1. [View alignment](#)

Reverse Maximised PSI-BLAST results:

Page 1 of 1. Total hits: 13 Selected: 0. [Select all](#) [Deselect all](#)

Cluster Detail	Code	Title	Organism	%ID	Query rgn.	Target rgn.	Filter	E value
♦	AAA35532.1	Not given	Hsapiens	99.0	1 - 516	1 - 516	1	0.00E+00
♦	AAF19049.1	Not given	Mmusculus	80.0	1 - 538	1 - 537	1	0.00E+00
♦	AAC53142.1	Not given	Mmusculus	90.0	1 - 538	1 - 538	1	0.00E+00
♦	AAC38001.1	ribonucleoprotein	Xlaevis	78.0	1 - 538	1 - 538	1	0.00E+00
♦	AAB81552.1	Not given	Hsapiens	100.0	1 - 189	1 - 189	1	1.00E-111
♦	CAA98241.1	Not given	Celegans	36.0	7 - 538	54 - 643	2	0.00E+00
♦	AAF10933.1	Not given	Dradiodurans	34.0	7 - 538	15 - 531	2	0.00E+00
♦	AAF43876.1	Not given	Dmelanogaster	17.0	1 - 537	64 - 639	3	1.00E-126
♦	CAB46406.1	hypothetical protein	Scorcolor	26.0	266 - 381	273 - 372	3	6.00E-25
♦	AAC53043.1	telomerase protein-1	Mmusculus	21.0	264 - 378	513 - 627	3	6.00E-05
♦	AAB51690.1	telomerase protein component 1	Rpovegicus	23.0	284 - 375	517 - 627	3	2.00E-04
♦	AAC46501.1	telomerase component p80	Tthermophila	27.0	264 - 329	361 - 425	3	7.00E-04
♦	AAC51107.1	telomerase-associated protein TP-1	Hsapiens	25.0	264 - 327	507 - 669	3	0.00E-03

GenBank SWISS-PROT PDB

## FIG. 23

AIEye output (January 2, 2001 1:29 PM)

36/40

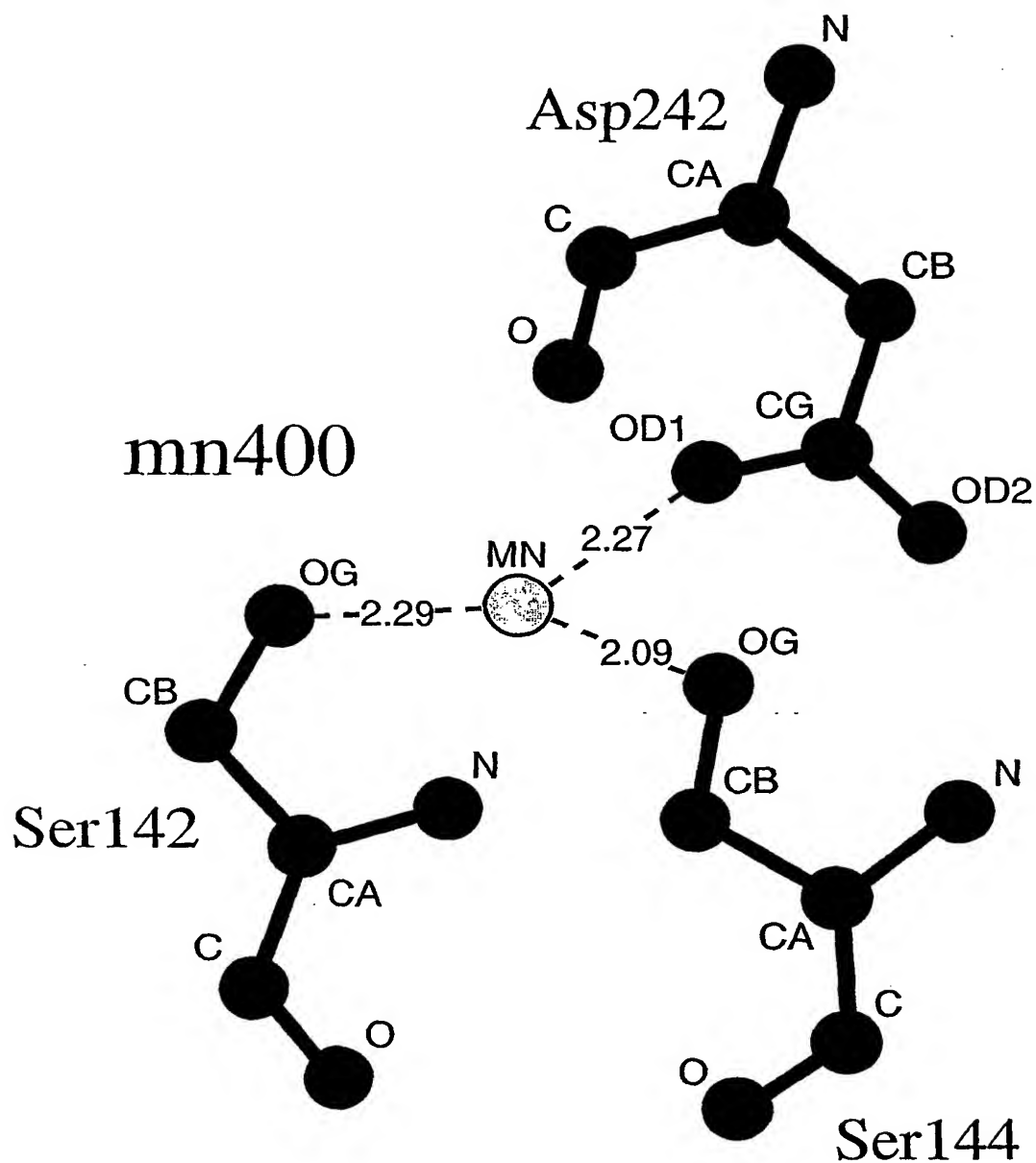
```

P101SS 10 20 30 40 50 60 70 80
1JUN MEESYNQMPLNEKQIANSDQGYVWQVTDNRLHRFLCFGSEGGTTYIKQKLGLENAEALIRLEDGRGCEVIQEKSF
-----
P101SS 90 100 110 120 130 140 150 160
1JUN SQEGRTTKQEPMLPALAICSQCSDISTKQAAFKAVSEVCRIPHTLFTFIQKKDLKESMKCGMWGRALRKAADWYNEKG
-----
P101SS 170 180 190 200 210 220 230 240
1JUN GMA LALAVTKYKQRNGWGHKDLLRLSHLKPSSGLAIVTKYITKGWKEVHELYKEKALSVEYETKLLKYLEAVEKVKRTKD
-----
P101SS 250 260 270 280 290 300 310 320
1JUN ELEVILIEEHLVREHLTNHLKSKEVWKA LLQEMPLTALLRLGKMTANSVLEPGNSEVSLYCEKLCNEKLLKKARIH
-----
P101SS 330 340 350 360 370 380 390 400
1JUN PFHILIALETYKTGHGLRGKLLKWRPDEEILKALDAAFYKTFKTVEPTGKRFL--L--AVDVSA SMNQRPVLG SILNASTVA
-----
P101SS 410 420 430 440 450 460 470 480
1JUN -----DSDIAFLDGGSGSII PHDFRRMKEFVSTV
-----
P101SS 490 500 510 520 530 540 550 560
1JUN A-----AMCMVVTRTEKDSYVVAFSDENVPCPVTTDMTLQQVLNAMSQI--PAGGTDCSLPMIWAQKTNT---PADVFIYF
NEQLKSKTLFSLNQYSEEFRIHFTFKEFQNNPNPRSLVKPITQLLGRHTATGIRKVVRELFNITNGARKNAFKILVVI
-----
P101SS 570 580 590 600
1JUN TONETFAGGVHPAIALREYRKKMDIPAKLIVCGMTSNGF-----TIADPPDRG-MLDMCGFDTGALDVI RNF TLOMI
TOGEKFGDPLGYEDVIPE-ADREGVI RYVIGVGDAFRSEKSRQELNTIASKPPRHDVYFQYNNFE--ALKTIQNQLREKIF
-----
P101SS 610 620 630 640
1JUN A

```

37/40

FIG. 24A



1jlm: MN 400

38/40

FIG. 24B





39/40

# FIG. 25

AlEye output (January 4, 2001 3:18 PM)

P10155	MEESVN	10	20	30	40	50	60	70
AAF19049.1	MEGSAN	10	20	30	40	50	60	70
AAC38001.1	MEATMD	10	20	30	40	50	60	70
CAA98241.1	MADELNEFOEAGNFNEEALMRLSNVCA	10	20	30	40	50	60	70
	RLRMOMLES	10	20	30	40	50	60	70
	SDVEITV	10	20	30	40	50	60	70
	VDGELKRV	10	20	30	40	50	60	70
	PRQMEKVK	10	20	30	40	50	60	70
	QDQVENN	10	20	30	40	50	60	70
	AGGFVFP	10	20	30	40	50	60	70
P10155	VTDMNRLH	80	90	100	110	120	130	140
AAF19049.1	FLCFGSEGG	80	90	100	110	120	130	140
AAC38001.1	TYI	80	90	100	110	120	130	140
CAA98241.1	KEQKLGLENAEALIRLIEDGRGCE	80	90	100	110	120	130	140
	VIQEI	80	90	100	110	120	130	140
	KSFSGEGRTTKQEPMLFAL	80	90	100	110	120	130	140
	ICSQCS	80	90	100	110	120	130	140
P10155	VTDMN	150	160	170	180	190	200	210
AAF19049.1	VRRFLCFGSEGG	150	160	170	180	190	200	210
AAC38001.1	TYI	150	160	170	180	190	200	210
CAA98241.1	KEQKLGLENAEALIRLIEDGRGCE	150	160	170	180	190	200	210
	VIQEI	150	160	170	180	190	200	210
	KSFSGEGRTTKQEPMLFAL	150	160	170	180	190	200	210
	ICSQCS	150	160	170	180	190	200	210
P10155	VSDMNRL	220	230	240	250	260	270	280
AAF19049.1	RRFLCFGSEGG	220	230	240	250	260	270	280
AAC38001.1	TYI	220	230	240	250	260	270	280
CAA98241.1	EEKKLGOENAEALLRLIEDGKGCE	220	230	240	250	260	270	280
	VQEI	220	230	240	250	260	270	280
	KTFSGEGRAAKQEPTL	220	230	240	250	260	270	280
	FALAVCSQCSD	220	230	240	250	260	270	280
P10155	VSDTQVRR	290	300	310	320	330	340	350
AAF19049.1	FLILGSDKGSYHQSEK	290	300	310	320	330	340	350
AAC38001.1	ITIDNAQRIKIIEQNGHMLKELAL	290	300	310	320	330	340	350
CAA98241.1	INAENRNPKMNAMIFTLAICAR	290	300	310	320	330	340	350
	IST	290	300	310	320	330	340	350
	STKQ	290	300	310	320	330	340	350
	INTKQ	290	300	310	320	330	340	350
	IKTKQ	290	300	310	320	330	340	350
	HDTTKTECPMLNAYSDYIRALHDS	290	300	310	320	330	340	350
	ADLPEVORTPTHLFEFVDYCQ	290	300	310	320	330	340	350
	TISESTKAGGAKSSTGWGRSMNA	290	300	310	320	330	340	350
	ISKWYT	290	300	310	320	330	340	350
P10155	EKGGMAL	360	370	380	390	400	410	420
AAF19049.1	ALAVTKYKQ	360	370	380	390	400	410	420
AAC38001.1	NRNGWSHKDLLRLSH	360	370	380	390	400	410	420
CAA98241.1	---LKPSSEGLA	360	370	380	390	400	410	420
	---VTKYITKG	360	370	380	390	400	410	420
	---WKEVHEL	360	370	380	390	400	410	420
P10155	EKGGMAL	430	440	450	460	470	480	490
AAF19049.1	VALVTKYKQ	430	440	450	460	470	480	490
AAC38001.1	NRNGWSHKDLLRLSH	430	440	450	460	470	480	490
CAA98241.1	---LKPSSEGLA	430	440	450	460	470	480	490
	---VTKYITKG	430	440	450	460	470	480	490
	---WKEVHEL	430	440	450	460	470	480	490
P10155	TKDALN	500	510	520	530	540	550	560
AAF19049.1	LAMAVTKYKQ	500	510	520	530	540	550	560
AAC38001.1	NRNGWSHKDLLRLSH	500	510	520	530	540	550	560
CAA98241.1	---LKPSSEGLA	500	510	520	530	540	550	560
	---VTKYITKG	500	510	520	530	540	550	560
	---WKEVHEL	500	510	520	530	540	550	560
P10155	TKTEKL	570	580	590	600	610	620	630
AAF19049.1	AMLTKYPQ	570	580	590	600	610	620	630
AAC38001.1	REGWSHRDLFRLAHPNLMDS	570	580	590	600	610	620	630
CAA98241.1	SRSHGQSEDRLE	570	580	590	600	610	620	630
	QQLFRFAVKGDLVKKRKM	570	580	590	600	610	620	630
	SVEEAEVEKV	570	580	590	600	610	620	630
P10155	YKEKAL	640	650	660	670	680	690	700
AAF19049.1	SV	640	650	660	670	680	690	700
AAC38001.1	---ETEKLLKYLEAVE	640	650	660	670	680	690	700
CAA98241.1	VEKVKRTKDELEVI	640	650	660	670	680	690	700
	HLIEEHRVREHLLTNHLKS	640	650	660	670	680	690	700
	KEVWKA	640	650	660	670	680	690	700
	LLQ-EMPLT	640	650	660	670	680	690	700
P10155	YKEKAL	710	720	730	740	750	760	770
AAF19049.1	SV	710	720	730	740	750	760	770
AAC38001.1	---EAEKLLKYLEAVE	710	720	730	740	750	760	770
CAA98241.1	VEKVKRTKDDLEVI	710	720	730	740	750	760	770
	HLIEEHRVREHLLTNHLKS	710	720	730	740	750	760	770
	KEVWKA	710	720	730	740	750	760	770
	LLQ-EMPLT	710	720	730	740	750	760	770
P10155	YKEKEL	780	790	800	810	820	830	840
AAF19049.1	SP	780	790	800	810	820	830	840
AAC38001.1	---ETEKVLKYLEATE	780	790	800	810	820	830	840
CAA98241.1	RVKRTKDELEI	780	790	800	810	820	830	840
	HLIDEYRLVREHLLTNHLKS	780	790	800	810	820	830	840
	KEIWKSLQ-DMPLT	780	790	800	810	820	830	840
P10155	WDKKAL	850	860	870	880	890	900	910
AAF19049.1	KLPYTEEQ	850	860	870	880	890	900	910
AAC38001.1	LIKEEQSRALNLVEAYLKL	850	860	870	880	890	900	910
CAA98241.1	NEQSEEVIAAIKKHGLV	850	860	870	880	890	900	910
	REHLPTTSLNSKL	850	860	870	880	890	900	910
	VWETLFDV	850	860	870	880	890	900	910
	SMPMT	850	860	870	880	890	900	910
P10155	ALLRNL	920	930	940	950	960	970	980
AAF19049.1	GKMTANSV	920	930	940	950	960	970	980
AAC38001.1	LEPGNSEVSL	920	930	940	950	960	970	980
CAA98241.1	CEKLSNEKLLKKARI	920	930	940	950	960	970	980
	HPFFHILIALETYK	920	930	940	950	960	970	980
	KGHLRGK	920	930	940	950	960	970	980
	LKWRPDEEIL	920	930	940	950	960	970	980
	KALDAAF	920	930	940	950	960	970	980
P10155	ALLRNL	990	1000	1010	1020	1030	1040	1050
AAF19049.1	GKMTANSV	990	1000	1010	1020	1030	1040	1050
AAC38001.1	LEPGNSEVSL	990	1000	1010	1020	1030	1040	1050
CAA98241.1	CEKLSNEKLLKKARI	990	1000	1010	1020	1030	1040	1050
	HPFFHILIALETYK	990	1000	1010	1020	1030	1040	1050
	KGHLRGK	990	1000	1010	1020	1030	1040	1050
	LKWRPDEEIL	990	1000	1010	1020	1030	1040	1050
	KALDAAF	990	1000	1010	1020	1030	1040	1050
P10155	ALLRNL	1060	1070	1080	1090	1100	1110	1120
AAF19049.1	GKMTANSV	1060	1070	1080	1090	1100	1110	1120
AAC38001.1	LEPGNSEVSL	1060	1070	1080	1090	1100	1110	1120
CAA98241.1	CEKLSNEKLLKKARI	1060	1070	1080	1090	1100	1110	1120
	HPFFHILIALETYK	1060	1070	1080	1090	1100	1110	1120
	KGHLRGK	1060	1070	1080	1090	1100	1110	1120
	LKWRPDEEIL	1060	1070	1080	1090	1100	1110	1120
	KALDAAF	1060	1070	1080	1090	1100	1110	1120
P10155	ALLRNL	1130	1140	1150	1160	1170	1180	1190
AAF19049.1	GKMTANSV	1130	1140	1150	1160	1170	1180	1190
AAC38001.1	LEPGNSEVSL	1130	1140	1150	1160	1170	1180	1190
CAA98241.1	CEKLSNEKLLKKARI	1130	1140	1150	1160	1170	1180	1190
	HPFFHILIALETYK	1130	1140	1150	1160	1170	1180	1190
	KGHLRGK	1130	1140	1150	1160	1170	1180	1190
	LKWRPDEEIL	1130	1140	1150	1160	1170	1180	1190
	KALDAAF	1130	1140	1150	1160	1170	1180	1190
P10155	ALLRNL	1200	1210	1220	1230	1240	1250	1260
AAF19049.1	GKMTANSV	1200	1210	1220	1230	1240	1250	1260
AAC38001.1	LEPGNSEVSL	1200	1210	1220	1230	1240	1250	1260
CAA98241.1	CEKLSNEKLLKKARI	1200	1210	1220	1230	1240	1250	1260
	HPFFHILIALETYK	1200	1210	1220	1230	1240	1250	1260
	KGHLRGK	1200	1210	1220	1230	1240	1250	1260
	LKWRPDEEIL	1200	1210	1220	1230	1240	1250	1260
	KALDAAF	1200	1210	1220	1230	1240	1250	1260
P10155	ALLRNL	1270	1280	1290	1300	1310	1320	1330
AAF19049.1	GKMTANSV	1270	1280	1290	1300	1310	1320	1330
AAC38001.1	LEPGNSEVSL	1270	1280	1290	1300	1310	1320	1330
CAA98241.1	CEKLSNEKLLKKARI	1270	1280	1290	1300	1310	1320	1330
	HPFFHILIALETYK	1270	1280	1290	1300	1310	1320	1330
	KGHLRGK	1270	1280	1290	1300	1310	1320	1330
	LKWRPDEEIL	1270	1280	1290	1300	1310	1320	1330
	KALDAAF	1270	1280	1290	1300	1310	1320	1330
P10155	ALLRNL	1340	1350	1360	1370	1380	1390	1400
AAF19049.1	GKMTANSV	1340	1350	1360	1370	1380	1390	1400
AAC38001.1	LEPGNSEVSL	1340	1350	1360	1370	1380	1390	1400
CAA98241.1	CEKLSNEKLLKKARI	1340	1350	1360	1370	1380	1390	1400
	HPFFHILIALETYK	1340	1350					

FIG. 25 (contd.)

AlEye output (January 4, 2001 3:18 PM)

